CONTRIBUTIONS TO STATISTICAL TECHNIQUES FOR TWO AND THREE DIMENSIONAL MEASUREMENT PROBLEMS

By

James Robert Lackritz

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1977

To my family and all other non-believers.

ACKNOWLEDGEMENTS

There are many people to whom I owe my sincere thanks for their help and encouragement during my graduate years at the University of Florida.

First and foremost, I will always be indebted to Dr. Richard Scheaffer for his inspiration and unbounded patience in helping me in my quest for my doctoral degree. He will always be remembered, not just for his statistical accomplishments, but as well for his great personal qualities as a human being.

I also wish to express my deepest gratitude to Dr. William Mendenhall for his helpful guidance during my graduate years.

I would like to thank the individual members of my committee for their help and interest in my studies.

Finally, I am grateful to Betty Rovira, for her patience and help in typing the final draft.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS iii

ABSTRACT vii

CHAPTER

1 INTRODUCTION 1

1.1 History of Quantitative Microscopy . . 1
1.2 Models for Random Division of Space . 2
1.3 Applications 4

2 BASIC STEREOLOGICAL RELATIONSHIPS 6

2.1 The Mosaic 6
2.2 Basic Stereological Results 7
2.3 The Spatial Covariance Function . . . 11
2.4 Poisson Lines Model 16
2.5 Meijering Cell Model 17
2.6 Boolean Scheme 18

3 LINEAL SAMPLING 21

3.1 Introduction 21
3.2 Lineal Sampling 22
3.3 Model Development 23
3.4 Asymptotic Distribution of the Estimator $L_j/\sqrt{L}$ 26
3.5 Basic Properties of Segment Counts . . 28
3.6 Asymptotic Distribution of $M_j/\sqrt{L}$ . . . 34
3.7 Joint Distribution of $\underline{M}/\sqrt{L}$ 37
3.8 Asymptotic Variance-Covariance Matrix of $\underline{M}/\sqrt{L}$ 39
3.9 Variance Approximations 42

TABLE OF CONTENTS - Continued

Page

CHAPTER

3.10 The Results of Hilliard and Cahn . . . 46
3.11 Poisson Lines Model 49
3.12 Meijering Cell Model 50
3.13 Tests on the Mosaic Using Segment Counts 51
3.14 Spatial Covariance Function Approach to Using Lineal Sampling: Poisson Process 54
3.15 Boolean Scheme 55

4 AREAL SAMPLING 59

4.1 Introduction 59
4.2 Asymptotic Distribution of $\underline{A}/\sqrt{\underline{A}}$ . . . 60
4.3 Estimation of the Variance Components. 62
4.4 Poisson Process 63
4.5 Meijering Cell Model 65
4.6 Comparisons with the Results of Hilliard and Cahn 65
4.7 Spatial Covariance Function Approach . 67

5 POINT SAMPLING 69

5.1 Introduction 69
5.2 Random Points on the Plane 69
5.3 Systematic Points on a Line 70
5.4 Estimation of the Variance 73
5.5 Covariance Between Colors 75
5.6 Asymptotic Distribution of the Estimates 76
5.7 Extensions of Point Sampling 85
5.8 Systematic Points on Random Lines . . 88
5.9 Comparisons with Hilliard and Cahn . . 89

6 MULTISTAGE SAMPLING 91

6.1 Introduction 91

TABLE OF CONTENTS - Continued

Page

CHAPTER

6.2 Lineal Subsampling 91
6.3 Point Subsampling 93

7 NUMERICAL RESULTS 96

7.1 Introduction 96
7.2 Results: Table I 97
7.3 Results: Table II 98
7.4 Results: Table III 100
7.5 Results: Table IV 100
7.6 Conclusions 101

BIBLIOGRAPHY 107

BIOGRAPHICAL SKETCH 108

Abstract of Dissertation Presented to the
Graduate Council of the University of Florida in Partial
Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

CONTRIBUTIONS TO STATISTICAL TECHNIQUES FOR TWO AND THREE DIMENSIONAL MEASUREMENT PROBLEMS

By

James Robert Lackritz

August 1977

Chairman: Dr. Richard L. Scheaffer
Major Department: Statistics

Estimation of structural properties of two and three-dimensional figures can be achieved through areal, lineal, and point sampling techniques. Such properties might include areal and volume fractions, surface area per unit volume, and boundary length per unit area. Unbiased estimators for the individual parameters are well-known, and have been used for quite some time in the research associated with this field. Additional properties of the individual estimators are developed, including consistency and asymptotic normality, and practical variance approximations are given. These properties are derived under the assumption of a general cell model for the construction of the population figure. Once the general results are obtained, special cases of more specific models are discussed.

The variance approximations for the estimators are obtained by making use of the available sampling information, so that the usual inferential techniques of interval estimation and hypothesis testing can be employed. Often, the assumption of a special model yields a simplified variance expression for the individual estimators.

A second approach to areal and volume-fraction analysis makes use of the spatial covariance function of the underlying model to generate a variance expression for the areal or volume-fraction estimator. The results from this approach are tied in with the results obtained from the assumption of the general cell model.

The final chapter assesses the use of numerical data to check the variance expressions derived in previous chapters. Four different figures were sampled, and the numerical results allow for the comparisons of the individual models.

CHAPTER 1

INTRODUCTION

1.1 History of Quantitative Microscopy

The problem of estimating structural properties of two and three-dimensional figures has been studied for over 100 years by engineers, mathematicians, and stereologists. DeHoff and Rhines [5] review the history of Quantitative Microscopy in summarizing the basic results that have been developed. The first work in this field dates back to 1848, when Delesse proved mathematically that the area occupied by each constituent of a cross-section of a rock is proportional to its volume in the mass of the rock. The first work in lineal analysis was introduced by Rosiwal in 1898, while in 1930, Thompson looked at measurements from point counts. Starting in 1945, more exact statistical relationships for two-dimensional sections were discovered in independent studies, first by Saltykov, and then by several scientists in the United States, including Rhines. The new ideas showed the direct proportionality between the number of intersections of a line with a phase on the figure and the boundary length of the phase of a two-dimensional feature or the surface area of the phase on a three-dimensional feature. Throughout the

the years, investigators have looked at the basic measurements that can be obtained through lineal, areal, and point sampling techniques, with a view towards developing estimators for the basic properties of the figure. Some important recent results in point sampling have been outlined by Koop [9], who studied 21 different methods of point sampling on the plane for purposes of areal fraction estimation. Under minimal assumptions, unbiased estimators have been generated by the researchers for such parameters as areal and volume-fractions, feature boundary length in two dimensions, and surface area in three dimensions. The culmination of the individual results is a mathematical n-dimensional theorem of Miles [15], from which many of the individual unbiased estimators can be derived. These results will be considered in more detail in Chapter 2.

1.2 Models for Random Division of Space

The work of Miles [15], generalizing the earlier individual results into one relationship, yields unbiased estimators for the two and three-dimensional properties with minimal assumptions on the underlying figure or the sampling scheme from which the measurements are taken. Many theoretical models for the random division of space have been developed to try to account for the construction of the mosaic. Several of these models will be discussed throughout the next four chapters.

One of the simplest and most theoretically investigated models is the Poisson Lines Model, whose properties have been developed and discussed by many scientists, including Miles [13] and Switzer [24]. This model allows for the division of space into a system of convex polygons by placing a series of random lines in a coordinate system, where the lines serve as boundaries of the individual cells. Miles [13] developed many important properties of the cells, including the behavior of a line transect that is randomly placed inside the cell configuration. Switzer [24] showed that the alternation among phases along a randomly placed line transect in the process is Markovian for a finite-phase Poisson Process.

The Meijering Cell Model, the result of which is sometimes known as Voronoi polygons, is discussed in detail by Gilbert [6], and again is further expanded on by Miles [14]. This model arises from the random placement of points to expand outward in all directions at a constant rate until meeting with another cell, upon which the boundary is defined. This model, which can be paralleled to a germination-growth model, thus results in a set of non-overlapping convex polygons.

The third model to be used will be the Boolean Scheme, discussed by Serra [22]. Like the Meijering Cell Model, this model starts with a random point process, but then random sets are placed around the points, allowing for set overlap, unlike either of the previous two models. Serra also looks at a different approach to volume-fraction estimation, which uses the spatial covariance function of the underlying model, if a particular model can be assumed.

The assumption of an underlying model for the mosaic gives a probability structure to the mosaic under consideration. Chapters 3 through 6 serve to tie the underlying probability structure to properties of the estimators, so that additional properties can be developed. Among the properties discussed will be consistency, asymptotic distributions, and variance approximations for the estimators. Hilliard and Cahn [8] developed variance approximations for areal-fraction estimators under certain assumptions on the mosaic, and their results will be shown to be a special case of the variance approximations obtained from more general models.

1.3 Applications

The major results discussed by DeHoff and Rhines [5] and generalized by Miles [15] have diverse applications into a wide range of fields, including materials science, ecology, physiology, and biology. DeHoff and Rhines [5] discuss how a structure of spherical copper powder may be examined under a microscope to estimate the fraction of the powder occupied by voids between the interparticle welds. An ecologist may be able to use an areal photograph of a piece of land to estimate the boundary length for a particular river or lake, and additional ecological examples are mentioned in detail by Pielou [16] and Matérn [12]. Weibel [25] presents a section on the utilization of electron micrographs of liver cells for useful information on cell sizes and orientation, especially relevant in cancer research and drug-related experiments. Lang and Melhuish [10] looked at estimating lengths and diameters of plant roots by passing a plane underground through the root system.

The applications are numerous, especially for the fields associated with the physical, agricultural, and ecological sciences. Many different useful measurements may be obtained through the various sampling techniques, not only for generating unbiased estimators for the parameters, but also for obtaining variance approximations for inferential purposes.

CHAPTER 2

BASIC STEREOLOGICAL RELATIONSHIPS

2.1 The Mosaic

Throughout the remaining chapters, the term mosaic will refer to a two-dimensional structure composed of two or more distinctive features called phases. The general topic under research involves the estimation of various structural properties of the mosaic through geometric sampling procedures, and the development of additional properties of these estimators, such as consistency, asymptotic distribution, and variance approximations. The details of interest on the mosaic might include:

(1) The areal fractions of the mosaic occupied by different phases, which in turn may estimate the volume-fractions if the mosaic represents a planar sectioning of a three-dimensional figure.

(2) Boundary length of certain two-dimensional features, which in turn, may estimate the corresponding surface area of three-dimensional features.

(3) Possible feature orientation or homogeneity on the mosaic.

The information needed to generate estimates for the parameters of the mosaic can be obtained through many different sampling procedures, three of which will be discussed in detail in the succeeding chapters. They are:

(1) Randomly selecting one or more areal sections from the mosaic.

(2) Randomly locating one or more linear transects within the boundaries of the mosaic.

(3) Randomly locating a set of points on the mosaic.

Estimation of structural properties of the mosaic through areal, lineal, and point sampling has a rich history in quantitative microscopy and other disciplines, as pointed out in Chapter 1. The remaining sections of this chapter will review some of the known results, which will be used to build upon in the remaining chapters.

2.2 Basic Stereological Results

Throughout the years, researchers in the field of quantitative microscopy have strived to develop unbiased estimators for the parameters of the mosaic. The population quantities of interest include volumes and surface areas of three-dimensional phases and areas and boundary lengths of two-dimensional phases. The sampling devices used are planes, lines, and points, so that the sample measurements arise from the intersection between the sampling device and the phase of the mosaic. The measurements from an areal section include the total area of the sample, A, and the area occupied by the j^{th} phase, denoted by A_j. The lineal transect measurements of interest are L, the length of the transect, L_j, the total length of the intersection of the line with the j^{th} phase, and I_j, the number of intersections of the transect with the boundaries of the j^{th} phase. Similarly for point sampling, the number of points, n, and the point count for phase j, N_j, will be important measurements to the experimenter.

A set of unbiased estimators for the parameters of the mosaic that were developed individually by the various scientists over the years were generalized by Miles [15] into a single result based on an exercise in n-dimensional geometrical probability. Miles' theorem states:

> To estimate a t-dimensional quantity of a d-dimensional mosaic using an s-dimensional sampling device yielding $(s + t - d)$-dimensional measurements, for s, t, and d non-negative integers such that $t \le d$ and $s \le d$, let
>
> D_d = the measure of the population figure
>
> D_s = the measure of the intersection of the sampling device with the population figure
>
> X_t = the measure of the feature of interest on the population figure
>
> X_{s+t-d} = the measure of the intersection of the sampling device with the feature of interest.
>
> (Here, measure is defined to be volume in three dimensions, area in two dimensions, length in one dimension, and point counts for zero dimensions.)

Suppose the sampling device is randomly located in the space and that D_s is not zero. Then

$$\frac{E(X_{s+t-d})}{E(D_s)} = k_d(s,t)\,\frac{X_t}{D_d} \qquad (2.2.1)$$

where

$$k_d(s,t) = \frac{\Gamma\left(\frac{s+1}{2}\right)\Gamma\left(\frac{t+1}{2}\right)}{\Gamma\left(\frac{s+t-d+1}{2}\right)\Gamma\left(\frac{d+1}{2}\right)} . \qquad (2.2.2)$$

Looking at the individual parameters of interest, the estimators obtained from this single relationship will be the same as those individually developed, as discussed in the history of quantitative microscopy by DeHoff and Rhines [5].

Suppose the parameter of interest is the volume-fraction occupied by a particular phase. Let V denote the volume of the three-dimensional figure, and V_j the volume of phase j on the figure. If an areal sample is used, then the intersection of the planar sample with phase j will be the phase j areas on the plane. With $d = t = 3$ and $s = 2$, $k_3(2,3) = 1$, so that $E(A_j)/E(A) = V_j/V$. Similarly, for a lineal transect, where $d = t = 1$, and for point sampling, where $d = t = 0$, one has

$$\frac{E(N_j)}{E(n)} = \frac{E(L_j)}{E(L)} = \frac{E(A_j)}{E(A)} = \frac{V_j}{V} . \qquad (2.2.3)$$

Also, for fixed areal fraction, A_j/A, when $d = t = 2$, lineal and point sampling will yield

$$\frac{E(N_j)}{E(n)} = \frac{E(L_j)}{E(L)} = \frac{A_j}{A} \; . \qquad (2.2.4)$$

It should be noted from (2.2.3) and (2.2.4) that in estimating a volume, areal, or lineal fraction with a lower-dimensioned sample-fraction, unbiased estimators result only when the intersection between the sampling device and the population figure of interest is constant. For example, if an areal sample is used, where A_j/A estimates V_j/V, and A is a non-degenerate random variable, then $E(A_j/A) \neq E(A_j)/E(A)$, so that A_j/A will be a biased estimator of V_j/V. In practice, this problem can be eliminated by predetermination of the sample "size," and locating the sampling scheme entirely within the borders of the population figure.

A second parameter of interest on a three-dimensional figure is the surface area, S_j, for a particular phase j. If the figure is sampled by a planar section of area A, the intersection of the plane with the phase will be the boundary length, B_j, of the phase on the plane. Therefore, with $d = 3$, $t = s = 2$,

$$k_3(2,2) = \frac{\Gamma(\frac{3}{2})\Gamma(\frac{3}{2})}{\Gamma(1)\Gamma(2)} = \frac{\pi}{4} \; ,$$

so that

$$\frac{4}{\pi}\frac{E(B_j)}{E(A)} = \frac{S_j}{V} . \qquad (2.2.5)$$

Similarly, if the boundary length per unit area for a phase j is to be estimated from a lineal transect, the sample measurements of interest are the number of intersections, I_j, the transect makes with the phase j boundary. With $d = 2$, $s = t = 1$,

$$k_2(1,1) = \frac{\Gamma(1)\Gamma(1)}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2})} = \frac{2}{\pi}$$

so that

$$\frac{\pi}{2}\frac{E(I_j)}{E(L)} = \frac{B_j}{A} , \qquad (2.2.6)$$

implying that twice the number of intersections per unit length of the transect with the phase boundary also serves as an unbiased estimator for the surface area per unit volume. Again, as mentioned previously, it will be essential that the denominator of the estimator be fixed to assure that the estimator is unbiased for the parameter. For a treatment of the biased estimation case, see Davy and Miles [4].

2.3 The Spatial Covariance Function

A large portion of Chapters 3, 4, and 5 will be devoted to getting variance approximations for the aforementioned estimators of Section 2.2. The notion of a spatial covariance function will be employed in some of this work.

Consider a two-dimensional mosaic, $\mathcal{G}$, of fixed overall size and any set β, of random size and location within the borders of $\mathcal{G}$.

Define

$$f(x) = \begin{cases} 1 & \text{if a point } x \in \mathcal{G} \text{ falls in set } \beta \\ 0 & \text{otherwise} \end{cases}$$

so that $P(x \in \beta) = E[f(x)]$. The expectation of $f(\cdot)$ can be found by first conditioning on β being of fixed size, and then taking the expectation, $E_\beta(\cdot)$, over the sizes of β. Thus,

$$\begin{aligned} E[f(x)] &= E_\beta\{E[f(x) \mid \beta \text{ fixed size}]\} \\ &= E_\beta\left\{\frac{1}{\text{mes } \mathcal{G}} \int_{\mathcal{G}} f(x)dx\right\} \\ &= \frac{E_\beta(\text{mes } \beta)}{\text{mes } \mathcal{G}} , \end{aligned}$$

where mes$(\cdot)$ denotes volume in three dimensions, area in two dimensions, length in one dimension, and point counts in zero dimensions. For the remainder of this section, let $p = E_\beta(\text{mes } \beta)/\text{mes } \mathcal{G}$.

For two points, x and $x+h$, a distance h apart and at an angle α from a fixed orientation, let

$$\begin{aligned} C(h) &= E[f(x)f(x+h)] = P[x, x+h \in \beta] \\ &= E_\beta\{E[f(x)f(x+h) \mid \beta \text{ fixed size}]\} \end{aligned}$$

$$= E_\beta\left\{\frac{1}{\text{mes } G}\int_G f(x)f(x+h)dx\right\}$$

$$= \frac{E_\beta(\text{mes } \beta \cap \beta_{-h})}{\text{mes } G} ,$$

where β_{-h} represents the set obtained by translating all points in β by a distance h in the direction opposite to α.

Note that

$$\text{(i)} \quad C(0) = p$$

$$\text{(ii)} \quad C(h) = C(-h)$$

$$\text{(iii)} \quad C(0) \geq C(h)$$

$$\text{(iv)} \quad C(\infty) = p^2.$$

The spatial covariance function, cov(h), is defined by

$$\text{cov}(h) = E[f(x)f(x+h)] - E[f(x)]E[f(x+h)]$$

$$= C(h) - p^2.$$

Furthermore, if β is a two-dimensional feature defined on the mosaic, an important result involving the covariance function is that the expected boundary length of β per unit area is equal to $-\pi \frac{d}{dh}\text{cov}(h)\Big]_{h=0} = -\pi\, C'(0)$.

If p is to be estimated, let the set S denote the sampling device on the mosaic, where S might represent a subarea, one or more line transects, or a set of points dropped on the mosaic.

Let

$$I_g = \int_G f(x)g(x)dx ,$$

where $g(x)$ is some weight function defined on the sampling scheme. In particular, $g(x)$ could be a simple function, defined by

$$g(x) = \frac{1}{\text{mes } S} \quad \text{if } x \in S,$$
$$= 0 \qquad \text{otherwise}$$

for which case

$$I_g = \frac{1}{\text{mes } S} \int_S f(x)dx .$$

Now,

$$\begin{aligned}
E(I_g^2) &= E\{[\int_G f(x)g(x)dx]^2\} \\
&= E[\int_G f(x)g(x)dx \int_G f(y)g(y)dy] \\
&= E[\int_G \int_G f(x)f(y)g(x)g(y)dx\, dy] \\
&= E[\int_G \int_G f(x)f(x+h)g(x)g(x+h)dx\, dh] \\
&= \int_G \int_G E[f(x)f(x+h)]g(x)g(x+h)dx\, dh \\
&= \int_G C(h) \int_G g(x)g(x+h)dx\, dh \\
&= \int C(h)G(h)dh ,
\end{aligned}$$

where

$$G(h) = \int_G g(x)g(x+h)dx$$

so that

$$\begin{aligned} V(I_g) &= E(I_g^2) - E^2(I_g) \\ &= \int_G C(h)G(h)dh - E[\int_G f(x)g(x)dx]E[\int_G f(y)g(y)dy] \\ &= \int_G C(h)G(h)dh - \int_G \int_G p^2 g(x)g(x+h)dx\, dh \\ &= \int_G cov(h)G(h)dh\,. \end{aligned} \tag{2.3.1}$$

Thus, knowledge of the covariance function for a particular structure can result in a variance expression for the area occupied by the random set β. Note that when

$$g(x) = \begin{cases} \dfrac{1}{mes\ S} & x \in S \\ 0 & \text{otherwise} \end{cases}$$

and the covariance function is stationary and isotropic, then for two random points x and y located on S

$$\begin{aligned} V(I_g) &= \frac{1}{(mes\ S)^2} \int_S \int_S cov(|y - x|)dx\, dy \\ &= \frac{1}{(mes\ S)^2} \int_0^\infty cov(h)b(h)dh\,, \end{aligned}$$

where b(h) is the probability density function of $h = |y - x|$.

With the knowledge of the spatial covariance function being necessary to obtain a variance expression for the areal-fraction estimator, a model for the construction of the mosaic would be very useful. The three models introduced in Chapter 1 will now be discussed in greater detail so that they may be used in the development of additional properties of the estimators.

2.4 Poisson Lines Model

As mentioned in Chapter 1, the Poisson Lines Model is one of the most common and widely investigated models for the random division of space. As defined by Switzer [24], this model allows for the random division of space by a set of random lines constructed as follows:

> Upon location of an origin, "0," at the center of the region, select n pairs of points $\{(r_i, \theta_i): i = 1,2,\ldots\}$ such that n has a Poisson distribution with parameter τ, and θ_i is uniform on the interval $(0,\pi)$. For each pair of points, fit the line
>
> $$x \cos \theta_i + y \sin \theta_i - r_i = 0.$$

This constitutes the realization of a Poisson Process with intensity parameter τ/π. The realization of this process leaves the area with a number of intersecting random lines, resulting in a set of convex polygons. Properties of this model have been discussed and developed by several mathematicians, including Switzer [24] and Miles [13]. Miles introduced some key results for areal fraction estimation:

(1) The points at which a line transect intersects the random lines form a Poisson Process of density $\lambda = 2\tau/\pi$.

(2) The distribution of the areas of the individual cells has mean

$$\mu = \frac{\pi}{\tau^2} \tag{2.4.1}$$

and variance

$$\sigma^2 = \frac{\pi^2(\pi^2 - 2)}{2\tau^4} \ . \tag{2.4.2}$$

2.5 Meijering Cell Model

The construction of the Meijering Cell Model, as discussed by Gilbert [6] and Miles [14], produces cells of convex polygons from an underlying Poisson Point Process of intensity τ. Each point expands outward at a constant rate until it meets with another cell, or equivalently, the cell assigned to a point, y_i, on the process contains all of the points that are closer to y_i than to any other point y_j $(j \neq i)$ of the process. For a two-dimensional model, Gilbert [6] showed that, with A_i denoting the area of the cell generated by y_i,

$$E(A_i) = 1.018\ \tau^{-\frac{2}{3}} \tag{2.5.1}$$

and

$$V(A_i) = .228\ \tau^{-\frac{4}{3}} \ . \tag{2.5.2}$$

2.6 Boolean Scheme

As developed by Serra [22], the Boolean Scheme is a model which allows for an overlap among cells.

Let $y_1, y_2, \ldots$ be the realization of a planar Poisson Point Process of intensity τ. Let $\beta_1, \beta_2, \ldots$ be a collection of independently and identically distributed (i.i.d.) random sets, with each β_i implanted at a corresponding y_i. The set $\beta = \bigcup_i \beta_i$ is called the realization of the Boolean Scheme, which will be known as phase one. The remainder of the planar mosaic is known as phase two. Serra notes that the Boolean Scheme "represents one of the first steps in modeling, when one admits negligible interaction between the particles [β_i]." This kind of model might be suggested for use in crystalline growth and land studies.

For estimation purposes, suppose the resulting mosaic is sampled by S, a fixed set located at random inside the the mosaic, $\mathcal{A}$, noting that S may represent a subarea of the mosaic, one or more lineal transects, or a set of points developed on the mosaic.

Suppose β_j is located at x. Then

$$\beta_j \oplus S = \{x : \beta_j \cap S \neq \phi\}.$$

Therefore,

$$P(S \cap \beta_j) = \frac{E \text{ mes } (\beta_j \oplus S)}{\text{mes } \mathcal{A}} .$$

Defining $N_\tau(S)$ to be the number of β_j sets that intersect with S, it can be shown that $N_\tau(S)$ has a Poisson distribution with parameter $\tau \, E \, \text{mes} \, (\beta \oplus S)$. For any point $x \in G$, define

$$k(x) = \begin{cases} 1 & \text{if } x \in \beta \\ 0 & \text{otherwise,} \end{cases}$$

and

$$K(h) = \int_G k(x)k(x+h)dx = \text{mes } (\beta \cap \beta_{-h}) .$$

If $S = \{x, x+h\}$, the set of two points differing by the vector h, then

$$\begin{aligned} \text{mes } (\beta \oplus S) &= 2 \, \text{mes} \beta - \text{mes } (\beta \cap \beta_{-h}) \\ &= 2K(0) - K(h) . \end{aligned}$$

Therefore,

$$\begin{aligned} C_2(h) &= P(x, x+h \in \beta^c) \\ &= P[N_\tau(S) = 0] \\ &= \exp[-\tau\{2K(0) - K(h)\}]. \end{aligned}$$

Defining

$$q = C_2(0) = P(x \in \beta^c) = e^{-\tau K(0)},$$

$$C_2(h) = q^2 e^{\tau K(h)} .$$

Also, with

$$P(x, x+h \in \beta^c) = P(x \in \beta^c) - P(x \in \beta^c , x+h \in \beta),$$

then

$$P(x \in \beta^c , x+h \in \beta) = q - q^2 e^{\tau K(h)},$$

so that

$$C_1(h) = P(x, x+h \in \beta) = 1 - 2q + q^2 e^{\tau K(h)} .$$

Therefore,

$$\begin{aligned} \mathrm{cov}(h) &= C_1(h) - p^2 = C_2(h) - q^2 \\ &= q^2 e^{\tau K(h)} - q^2 . \end{aligned} \qquad (2.6.1)$$

The estimation of p and q can now be achieved through the use of the covariance function technique discussed in Section 2.3.

With models available for the construction of the mosaic, the following chapters will serve to not only further develop the properties of Miles' estimators, but also to integrate these results with the available models for a total culmination of results available under varying conditions and assumptions on the mosaic.

CHAPTER 3

LINEAL SAMPLING

3.1 Introduction

Consider a two-dimensional mosaic composed of two or more phases. This mosaic could arise as a cross section of a three-dimensional figure or as a subsection of a larger planar figure. The purpose of this chapter is to investigate properties of estimators of planar features such as areas and boundary lengths, in the case of sampling by line transects. An areal photograph of a section of land may provide an environmentalist with estimates for the fraction of the land occupied by different "phases," such as trees or water reservoirs, along with possible additional information such as a boundary length or the length of a perimeter around a lake. A metallurgist may be interested in estimating the volume-fraction of the void in a piece of metal, while a medical technologist might examine a cross-section of a liver to estimate the fraction of the wall occupied by a certain type of cell.

The two properties of the mosaic to be studied in detail are:

(i) The areal fraction for each of the k phases, which may in turn estimate the corresponding volume-fractions of the phases, if the mosaic arises from a three-dimensional figure.

(ii) The boundary length for a given feature, which in turn is related to the surface area of a corresponding three-dimensional feature.

Estimation of these properties will be studied through lineal, areal, and point sampling techniques, the first of which is studied in this chapter.

3.2 Lineal Sampling

This section discusses the sampling procedure of randomly locating one or more transects of fixed length entirely within the borders of the mosaic, from which at least two sets of measurements will be obtained. The first will be the fractions of the transect passing through each of the k phases, which will be used to estimate the corrsponding areal fractions for each of the k phases. Another measurement of interest will be the segment counts on the transect for each of the k-phases, which will be useful not only in estimating the boundary length for each of the phases but also will contribute to approximating the variance of the areal-fraction estimators. The segment counts also will be used as a basis of a test for preferred orientation for features on the mosaic.

3.3 Model Development

The first step of the estimation process is the development of some asymptotic properties of the areal-fraction estimators. This will be done through the assumption of a model for the construction of the mosaic. The assumed model considers the mosaic as a union of a set of non-overlapping cells with cell sizes having finite second moments. Each cell is independently assigned to phase j with probability $p_j > 0$ for $j = 1,2,\ldots,k$ and $\sum_{j=1}^{k} p_j = 1$. Thus, a feature may consist of one or more adjoining cells of the same phase. A sampling transect of length L consists of one or more segments, each segment being composed of a random number of intersections with the original cells. Numerous phases may appear on an individual transect, but there is no guarantee that all k phases will show up on the transect.

Recall from Section 2.4 that both the Poisson Lines Model and the Meijering Cell Model result in cells with finite second moments, and hence, could result in a mosaic as discussed above. In practice, cells may be distinguishable in some cases, such as in observing the cross-section of the wall of the liver. However, in most cases where such a model may be employed, the individual cells may not be distinguishable. Also, many mosaics may not even have an underlying cell structure, especially when a large continuous feature such as a lake or river is studied, but the cell model will still be assumed in order to develop asymptotic properties of the estimators. The three specific models for cell development from Chapter 2 will be discussed later in more detail, but first the results will be derived in terms of a general cell model. A few terms will now be defined.

Let L equal the length of the sample transect randomly located on the mosaic, and let X_i denote the intercept length of the i^{th} cell cut by the transect. It is assumed that the cell intercepts cut by the transect have lengths that are independently and identically distributed with finite mean and variance, μ and σ^2 respectively. (This is indeed true for the Poisson Lines Model.)

Let

$$\phi_{ij} = \begin{cases} 1 & \text{if the } i^{th} \text{ cell cut by the transect belongs to phase } j \\ 0 & \text{otherwise.} \end{cases}$$

and let

N^* = the number of intersections between the transect of length L and the cell boundaries.

Generally, the value for N^* will not be obtainable from the data, unless the individual cells are distinguishable. Estimates for the expected value of N^*, denoted by $E(N^*)$, will be discussed later in this chapter.

Defining L_j by

$$L_j = \sum_{i=1}^{N^*} \phi_{ij} X_i \ , \qquad (3.3.1)$$

L_j will denote the total length of the intercepts of phase j on the transect of length L. Note that $\{\phi_{ij}X_i\}$ are i.i.d. with mean $p_j\mu$ and variance $p_j(1-p_j)\mu^2 + p_j\sigma^2$.

The cell intercepts on the transects can be viewed as a realization of a renewal process by considering the intersection points as the times of occurrences of different events (phase assignment). A standard result of renewal theory is that the average number of events per unit time converges to the inverse of the average length of time between events, or equivalently,

$$\frac{E(N^*)}{L} \to \frac{1}{\mu} \qquad (3.3.2)$$

as $L \to \infty$, assuming $\mu > 0$.

In looking at the limit of $E(L_j/L)$, Wald's Theorem [23] can be used since

$$E(\phi_{ij}X_i) = p_j\mu < \infty.$$

Therefore,

$$\lim_{L\to\infty} E(L_j/L) = p_j \,, \qquad (3.3.3)$$

implying that L_j/L is an asymptotically unbiased estimator of p_j. Also, by the Weak Law of Large Numbers and the fact that

$$N^*/L \xrightarrow{P} \frac{1}{\mu} \,,$$

L_j/L is a consistent estimator for p_j as $L \to \infty$.

The underlying mosaic is considered to be infinite in size, so that a sampling transect can intersect a large number of cells. This leads to an additional consideration on the mosaic. Since the cells have equal expected cell size, the total areal fraction, F_j, for the j^{th} phase on the mosaic, will be approximately equal to p_j. Thus, L_j/L can be thought of as a consistent and unbiased estimator of F_j.

3.4 Asymptotic Distribution of the Estimators $L_j/\sqrt{L}$

Define the vectors $\underline{L}$ and $\underline{p}$ by

$$\underline{L}' = (L_1 \; L_2 \cdots L_k)$$

$$\underline{p}' = (p_1 \; p_2 \cdots p_k) \; .$$

The next step in the estimation procedure is to derive the asymptotic distribution of $\frac{1}{\sqrt{L}}(\underline{L})$, which will allow for the usual inferential techniques of hypothesis testing and interval estimation. Point estimation has already accounted for $\frac{1}{L}\underline{L}$ being a consistent and asymptotically unbiased estimator of $\underline{p}$. The theory behind cumulative processes yields the theorems which give the joint asymptotic distribution of $\frac{1}{\sqrt{L}}(\underline{L} - \underline{p})$. To follow the definition of a cumulative process as stated by Smith [23], $\sum_{i=1}^{n} \phi_{ij} X_i$ is defined to be a cumulative process if:

(i) For fixed j, $\phi_{ij} X_i$ is a sequence of i.i.d. random variables for $i = 1,2,\ldots$

(ii) $\sum_{i=1}^{N^*} \phi_{ij} X_i$ is, with probability one, of bounded variation over every finite interval on the line.

The first condition holds by assumptions already stated. The second holds since N^* is finite with probability one on any finite interval, and $\phi_{ij} X_i$ is non-negative and finite.

The asymptotic distribution of $\frac{1}{\sqrt{L}}(\underline{L} - \underline{p})$ now can be derived as a result of Theorem 10 of Smith [23]:

> If $\sum_{i=1}^{n} (\phi_{i1} X_i \ \phi_{i2} X_i \ \cdots \ \phi_{ik} X_i)$ are k cumulative processes and $\sigma^2 < \infty$, then the vector $\sqrt{\mu/L}(\underline{L} - \underline{p})$ will converge in distribution as $L \to \infty$ to a multivariate normal random variable with zero mean vector and variance matrix B, where b_{st}, the stth element of B is given by
>
> $$b_{st} = \mathrm{cov}(\phi_{is} X_i - p_s X_i \, , \, \phi_{it} X_i - p_t X_i). \qquad (3.4.1)$$

To obtain the individual terms of the asymptotic variance-covariance matrix,

$$\begin{aligned} b_{jj} &= E[(\phi_{ij} - p_j)^2 X_i^2] \\ &= E[(\phi_{ij} - p_j)^2] E(X_i^2) \end{aligned}$$

since X_i and ϕ_{ij} are independent. Therefore

$$b_{jj} = p_j(1 - p_j)(\sigma^2 + \mu^2). \qquad (3.4.2)$$

For $s \neq t$, the covariance term,

$$\begin{aligned} b_{st} &= E[(\phi_{is} - p_s)(\phi_{it} - p_t)X_i^2] \\ &= E[(\phi_{is} - p_s)(\phi_{it} - p_t)]E(X_i^2) \\ &= -p_s p_t(\sigma^2 + \mu^2) . \end{aligned} \tag{3.4.3}$$

Therefore, as L gets large, the variance of L_j/L can be approximated from the above asymptotic variance. This term, which will be denoted as v_{jj}^*, can be written as

$$v_{jj}^* = \frac{p_j(1 - p_j)(\sigma^2 + \mu^2)}{L\mu} = \frac{p_j(1 - p_j)(1 + \frac{\sigma^2}{\mu^2})}{E(N^*)} . \tag{3.4.4}$$

The corresponding covariance term can now be approximated by

$$v_{st}^* = \frac{-p_s p_t(1 + \frac{\sigma^2}{\mu^2})}{E(N^*)} \tag{3.4.5}$$

for $s \neq t$. Now, the regular inferential procedures of hypothesis testing can be employed on the fractions $p_1, \ldots, p_k$, once estimates can be made for μ, σ^2, and $E(N^*)$.

3.5 Basic Properties of Segment Counts

With the establishment of the variance-covariance approximations for the lineal estimates, the next step is to try to estimate the modeling parameters of (3.4.4) and (3.4.5). Note that in using L_j/L to estimate p_j, μ and σ^2 are parameters of the mosaic, and will have to be estimated to get a reasonable approximation for v_{jj}^* and

V^*_{st}. Additional information that can be gathered from the sampling transect may be useful for this purpose. It is assumed that the individual X_i's cannot be measured. Instead, the transect appears as a series of segments of different phases. Each phase j segment is composed of a random number of cell intercepts, that random number having a geometric distribution with parameter p_j. The segment counts for each phase not only provide useful information in estimating the V^* matrix, but also allow for boundary length estimation of individual phases, along with a test for orientation and homogeneity of features on the mosaic. Define

$$f_{ij} = \begin{cases} 1 & \text{if the } (i-1)^{st} \text{ cell intercepted by the transect belongs to phase } j \text{ and the } i\text{-th cell intercepted is not a phase } j \text{ cell} \\ 0 & \text{otherwise} \end{cases}$$

and let $M_j = \sum_{i=2}^{N^*} f_{ij}$ for $j = 1,2,\ldots,k$. M_j will represent the number of individual segments of the j^{th} phase on the transect. For a mosaic with a large number of cells, the M_j's can be used to develop the theory behind the practical use of the segment counts in variance approximations, boundary length estimates, and orientation tests. Recall from (2.2.6) that $\pi/2$ times the number of intersections per unit length that the transect makes with the cell boundaries gives an unbiased estimate of the boundary length per unit area for the phase j features, with a corresponding result for surface area in three dimensions. In defining I_j to be the number of intersection points the transect makes with the boundaries of phase j features, $I_j = 2M_j$ unless the test line begins or ends with a phase j region. In any case, $2M_j$ gives a reasonable large sample approximation of I_j, so that any asymptotic results for I_j come about by looking at corresponding results for M_j.

Define

$$J_n = \text{the phase of the } n^{th} \text{ cell cut by the transect}$$

and

$$F_{vj}(x) = P(X_n \leq x \mid J_{n-1} = v,\ J_n = j).$$

The vector $\underline{M}' = (M_1\ M_2 \cdots M_k)$ constitutes a vector of functions defined on a Markov Renewal Process by definition of Pyke [17], with transition matrix

$$\{p_{vj}\} = \{p_j\}\ ,$$

since cells are independently shaded, and sojourn times governed by, $\{F_{vj}(x)\}$. The joint asymptotic distribution of $\frac{1}{\sqrt{L}}\underline{M}$ will be derived through the use of limit theorems developed by Pyke and Schaufele [18]. To equate the terminology of Markov Renewal Processes with the sampling results from the lineal transects, a few additional terms need to be defined. Let

$$\delta_{vj} = \begin{cases} 1 & \text{if } v = j \\ 0 & \text{otherwise,} \end{cases}$$

$$m_v^{(j)} = \text{(expected number of phase } v \text{ cells crossed by the transect between two consecutive segments of phase } j) + \delta_{vj}$$

$$= p_v/p_j \ ,$$

$$\eta_v = \sum_{j=1}^{k} p_{vj}\mu = \mu \ ,$$

$$\mu_{jj} = \text{expected return distance to phase } j \text{ on the transect}$$

$$= \sum_v m_v^{(j)} \eta_v$$

$$= \mu/p_j \ ,$$

and

$${}_vM_{wr} = \text{expected number of cells of phase } r \text{ on the line between a segment of phase } w \text{ until the next segment of phase } v$$

$$= \frac{p_r}{p_v}(1 - \delta_{vv}) + \delta_{wr} \ .$$

The theorems needed to derive the asymptotic distribution of $\underline{M}$ are all stated in terms of the moments of f_{ij}, so the next step is to determine these values. Without loss of generality, the attention will be focused on phase 1, since the procedures used to look at the other phases will be identical. For any other phase j, all subscripts for phase one can be replaced by the pertinent phase j, for $j = 2,3,\ldots,k$. To look at the moments of f_{i1}, let

$$\xi_{jm}(f_{i1}) = \int_0^\infty f_{i1}p_{jm}dF_{jm}(X_i)$$

$$= \begin{cases} p_m & \text{if } j = 1,\ m \neq 1 \\ 0 & \text{otherwise ,} \end{cases}$$

$$\xi_{jm}^{(2)}(f_{i1}) = \int_0^\infty f_{i1}^2 p_{jm}dF_{jm}(X_i)$$

$$= \begin{cases} p_m & \text{if } j = 1,\ m \neq 1 \\ 0 & \text{otherwise ,} \end{cases}$$

$$\zeta_j(f_{i1}) = \sum_{m=1}^{k} \xi_{jm}(f_{i1})$$

$$= \begin{cases} 0 & j \neq 1 \\ 1 - p_1 & j = 1, \end{cases}$$

and

$$\zeta_j^{(2)}(f_{i1}) = \sum_{m=1}^{k} \xi_{jm}^{(2)}(f_{i1})$$

$$= \begin{cases} 0 & j \neq 1 \\ 1 - p_1 & j = 1\ . \end{cases}$$

Using the Strong Law of Large Numbers given by Pyke and Schaufele for functions defined on a Markov Renewal Process,

$$\frac{M_1}{L} \xrightarrow{\text{a.s.}} \frac{\sum_{j=1}^{k} m_j^{(1)} \zeta_j(f_{i1})}{\mu_{11}} . \tag{3.5.1}$$

Since

$$\zeta_j(f_{i1}) = \begin{cases} 0 & \text{if } j \neq 1 \\ 1 - p_1 & j = 1 \end{cases}$$

and

$$m_1^{(1)} = 1 ,$$

$$\frac{M_1}{L} \xrightarrow{\text{a.s.}} \frac{p_1(1 - p_1)}{\mu} \tag{3.5.2}$$

as $L \to \infty$. Therefore, as $L \to \infty$,

$$\frac{M_j}{L} \xrightarrow{\text{a.s.}} \frac{p_j(1 - p_j)}{\mu} \tag{3.5.3}$$

for $j = 1,2,...,k$.

Through a Central Limit Theorem result of Pyke and Schaufele [18] for functions defined on a Markov Renewal Process, it will be shown that each M_j, suitably normed, will have an asymptotic normal distribution, and that the joint asymptotic distribution of the vector $\underline{M}$, suitably normed, will be multivariate normal with a closed-form expression for the variance-covariance matrix.

3.6 Asymptotic Distribution of $M_j/\sqrt{L}$

From Theorem 7.1(b) and Lemma 4.2 of Pyke and Schaufele [18],

$$\frac{1}{\sqrt{L}} \sum_{i=2}^{N^*} \left[f_{ij} - \frac{X_i p_j (1 - p_j)}{\mu} \right] \qquad (3.6.1)$$

will converge in distribution as L → ∞ to a standard normal random variable, provided the variance of each term in the sum (3.6.1) is finite. Recalling that

$$M_j = \sum_{n=2}^{N^*} f_{ij} \; ,$$

and defining

$$C_j = \frac{p_j (1 - p_j)}{\mu} \; ,$$

(3.6.1) becomes

$$\sqrt{L} \left(\frac{M_j}{L} - C_j \sum_{i=2}^{N^*} \frac{X_i}{L} \right) .$$

Now, standard renewal theory results give

$$\sqrt{L} \left(\sum_{i=2}^{N^*} \frac{X_i}{L} - 1 \right) \xrightarrow{P} 0$$

as L → ∞. Thus, working with (3.6.1) ultimately yields the asymptotic normality of $\sqrt{L}\left(\frac{M_j}{L} - C_j \right)$.

Again, without loss of generality, focus attention on phase 1, and define

$$h_{i1} = f_{i1} - X_i C_1$$

Lemma 4.2 of Pyke and Schaufele [18] yields the asymptotic variance expression for (3.6.1) as

$$\begin{aligned} B_1 = \mu_{11}^{-1} \Big\{ & \sum_j \zeta_j^{(2)}(h_{i1}) m_j^{(1)} \\ & + 2\sum_j \sum_{m\neq 1} \sum_{r\neq 1} \left[\xi_{jm}(h_{i1}) \zeta_r(h_{i1}) m_j^{(1)}\right] ({}_1M_{mr}) \Big\} \end{aligned} \qquad (3.6.2)$$

Following the previous definition of terms,

$$\begin{aligned} \xi_{jm}(h_{i1}) &= \int h_{i1}(X) p_{jm} dF_{jm} \\ &= \xi_{jm}(f_{i1}) - C_1 \mu p_m \\ &= \begin{cases} p_m - \mu C_1 p_m & j = 1,\ m \neq 1 \\ -\mu C_1 p_m & \text{otherwise,} \end{cases} \end{aligned} \qquad (3.6.3)$$

$$\begin{aligned} \xi_{jm}^{(2)}(h_{i1}) &= \int [f_{i1} - X_i A_1]^2 p_{jm} dF_{jm} \\ &= \xi_{jm}^{(2)}(f_{i1}) - 2C_1 \mu \xi_{jm}(f_{i1}) + C_1^2 p_m \mu^{(2)} \end{aligned}$$

where $\mu^{(2)} = \sigma^2 + \mu^2$. Therefore

$$\xi_{jm}^{(2)}(h_{i1}) = \begin{cases} p_m - 2C_1 \mu p_m + C_1^2 p_m \mu^{(2)} & j = 1,\ m \neq 1 \\ C_1^2 p_m \mu^{(2)} & \text{otherwise ,} \end{cases} \qquad (3.6.4)$$

$$\zeta_j(h_{i1}) = \sum_m \xi_{jm}(h_{i1})$$

$$= \begin{cases} -\mu C_1 & j \neq 1 \\ -p_1 - \mu C_1 & j = 1 \end{cases} \tag{3.6.5}$$

and

$$\zeta_j^{(2)}(h_{i1}) = \sum_m \xi_{jm}^{(2)}(h_{i1})$$

$$= \begin{cases} C_1\mu^{(2)} & j \neq 1 \\ (1 - p_1) - 2C_1\mu(1 - p_1) + C_1^2\mu^{(2)} & j = 1. \end{cases} \tag{3.6.6}$$

Now,

$$\sum_j m_j^{(1)}\zeta_j^{(2)}(h_{i1}) = 1 - p_1 - 2C_1\mu(1 - p_1) + \sum_j \frac{p_j}{p_1} C_1^2 \mu^{(2)}$$

$$= 1 - p_1 - 2C_1\mu(1 - p_1) + \frac{C_1^2\mu^{(2)}}{p_1} \tag{3.6.7}$$

and

$$2\sum_j \sum_{m\neq 1} \sum_{r\neq 1} [\xi_{jm}(h_{i1})\zeta_r(h_{i1})m_j^{(1)}]_1 M_{mr}$$

$$= 2\sum_{m\neq 1} \sum_{r\neq 1} (p_m - \mu C_1 p_m)(-\mu C_1)\left[\frac{p_r}{p_1}(1 - \delta_{r1}) + \delta_{mr}\right]$$

$$+ \sum_{j\neq 1} \sum_{m\neq 1} \sum_{r\neq 1} (-\mu C_1 p_m)(-\mu C_1)\frac{p_j}{p_1}\left[\frac{p_r}{p_1}(1 - \delta_{r1}) + \delta_{mr}\right],$$

which reduces to

$$\frac{-2\mu C_1}{p_1}\left[1 - p_1 - \mu C_1 \frac{(1 - p_1)}{p_1}\right] . \qquad (3.6.8)$$

Substituting into (3.6.2),

$$B_h = \frac{p_1}{\mu}\left\{1 - p_1 - 2C_1\mu(1-p_1) + \frac{C_1^2\mu^{(2)}}{p_1} - \frac{2\mu C_1}{p_1}\left[1 - p_1 - \frac{\mu C_1(1-p_1)}{p_1}\right]\right\}$$

$$= C_1[p_1^2 + (1 - p_1)^2] + C_1^2\left(\frac{\mu^{(2)}}{\mu} - 2\mu\right) \qquad (3.6.9)$$

Therefore, the asymptotic variance

$$V^*(M_j/\sqrt{L}) = C_j[p_j^2 + (1 - p_j)^2] + C_j^2\left(\frac{\mu^{(2)}}{\mu} - 2\right) , \qquad (3.6.10)$$

for $j = 1,2,\ldots,k$. Since $\mu^{(2)} < \infty$, $V^*(M_j/\sqrt{L}) < \infty$, and therefore, the asymptotic distribution of each $M_j/\sqrt{L}$ is normal, with mean C_j and variance $C_j[p_j^2 + (1 - p_j)^2] + C_j^2\left(\frac{\mu^{(2)}}{\mu} - 2\mu\right)$.

3.7 Joint Distribution of $\frac{1}{\sqrt{L}}\underline{M}$

With the univariate asymptotic normality of each $M_j/\sqrt{L}$ established, the next step is to prove the joint normality of $\underline{M}/\sqrt{L}$. The argument used to achieve the final result is given by Smith [23].

Recalling $h_{ij} = f_{ij} - X_iC_j$, write

$$\underline{h}_i = (h_{i1} \cdots h_{ik})$$

for $i = 2,3,\ldots$. Now $E(h_{ij}) = 0$ for all i and j, and let $\mathrm{cov}(h_{is},h_{it}) = a_{st}$ for $s \neq t$. Note that the a_{st} terms have not yet been developed, but will be used in finding the asymptotic covariance between $M_s/\sqrt{L}$ and $M_t/\sqrt{L}$. The term, $T_j = \sum_{i=2}^{N^*} h_{ij}/\sqrt{L}$ will converge in distribution to a normal random variable with mean zero. For any real vector $\underline{\theta} = (\theta_1\ \theta_2 \cdots \theta_k)$,

$$\underline{\theta}'\underline{T} = \sum_{j=1}^{k} \theta_j T_j = \sum_{j=1}^{k} \theta_j \sum_{i=2}^{N^*} \frac{h_{ij}}{\sqrt{L}}$$

$$= \frac{1}{\sqrt{L}} \sum_{i=2}^{N^*} \sum_{j=1}^{k} \theta_j h_{ij}$$

$$= \frac{1}{\sqrt{L}} \sum_{i=2}^{N^*} (\theta_1 h_{i1} + \theta_2 h_{i2} + \cdots + \theta_k h_{ik})$$

Letting

$$W_i = \frac{1}{\sqrt{L}}(\theta_1 h_{i1} + \theta_2 h_{i2} + \cdots + \theta_k h_{ik})$$

$$E(W_i) = 0 ,$$

and as $L \to \infty$,

$$V(W_i) = \frac{1}{L}\left[\sum_{j=1}^{k} \theta_j^2 V(h_{ij}) + 2 \sum_{s \neq t}\sum \theta_s \theta_t \text{cov}(h_{is}, h_{it})\right]$$

$$= \underline{\theta}'\underline{V}^*\underline{\theta} .$$

Since $\underline{\theta}'\underline{T}$ converges in distribution as $L \to \infty$ to a normal random variable with mean zero, one can see by the use of characteristic functions that $\underline{T}$ will have a multivariate normal distribution with zero mean vector. That is,

$$\sqrt{L}(\underline{M}/L - \underline{C}) \xrightarrow{D} N(\underline{0}, V^*).$$

3.8 Asymptotic Variance-Covariance Matrix of $\underline{M}/\sqrt{L}$

The covariance terms of $\underline{M}/\sqrt{L}$ can be obtained through Lemma 4.3 of Pyke and Schaufele [18], which yields an expression B_{st} similar to the B_h introduced earlier, except that when $s \neq t$, the expression deals with the covariances of h_{is} and h_{it} functions instead of the individual variances of the h_{ij} terms. Formally, let

$$B_{st} = \sum_j \zeta_j(h_{is}, h_{it}) m_j + \sum_j \sum_{m \neq 1} \sum_{r \neq 1} [\xi_{jm}(h_{is}) \zeta_r(h_{it})$$
$$+ \xi_{jm}(h_{it}) \zeta_r(h_{is})] m_j^{(1)} {}_1M_{mr} \qquad (3.8.1)$$

where

$$\xi_{jm}(h_{is}, h_{it}) = \int h_{is} h_{it} p_{jm} dF_{jm}$$
$$= -\int X_i [f_{is} C_t + f_{it} C_s] p_{jm} dF_{jm}$$
$$+ \int X_i^2 C_s C_t p_{jm} dF_{jm}$$

$$= \begin{cases} \mu^{(2)} C_s C_t p_m & j \neq s,t \text{ or } j = m = s \\ & \phantom{j \neq s,t \text{ or }} j = m = t \\ \mu^{(2)} C_s C_t p_m - \mu C_s p_m & j = s,\ m \neq s \\ \mu^{(2)} C_s C_t p_m - \mu C_t p_m & j = t,\ m \neq t \end{cases} \qquad (3.8.2)$$

and

$$\zeta_j(h_{is},h_{it}) = \sum_m \xi_{jm}(h_{is},h_{it})$$

$$= \begin{cases} \mu^{(2)}C_sC_t & j \neq s,t \\ \mu^{(2)}C_sC_t - \mu C_s(1-p_s) & j = s \\ \mu^{(2)}C_sC_t - \mu C_t(1-p_t) & j = t \end{cases} \qquad (3.8.3)$$

so that

$$\sum_j \zeta_j(h_{is},h_{it})m_j$$

$$= \sum_j \mu^{(2)}C_sC_t \frac{p_j}{p_1} - \mu C_s(1-p_s)\frac{p_s}{p_1} - \mu C_t(1-p_t)\frac{p_t}{p_1}$$

$$= \frac{1}{p_1}[\mu^{(2)}C_sC_t - \mu C_sp_s(1-p_s) - \mu C_tp_t(1-p_t)]. \quad (3.8.4)$$

Now,

$$\sum_j \sum_{m\neq 1} \sum_{r\neq 1} [\xi_{jm}(h_{is})\zeta_r(h_{it}) + \xi_{jm}(h_{it})\zeta_r(h_{is})]m_{j\,1}^{(1)}M_{mr}$$

$$= \sum_j \sum_{m\neq 1} \sum_{r\neq 1} [\{-\mu C_sp_m + p_m\delta_{js}(1-\delta_{ms})\}\{-\mu C_t + (1-p_t)\delta_{rt}\}$$

$$+ \{-\mu C_tp_m + p_m\delta_{jt}(1-\delta_{mt})\}\{-\mu C_s + (1-p_s)\delta_{rs}\}$$

$$\frac{p_j}{p_1}\left[\frac{p_r}{p_1}(1-\delta_{r1}) + \delta_{mr}\right]$$

$$= \sum_{j} \sum_{m\neq 1} \sum_{r\neq 1} [\mu^2 C_s C_t p_m - \mu C_s p_m (1 - p_t)\delta_{rt} - \mu C_t p_m \delta_{js}(1 - \delta_{ms})$$

$$+ p_m (1 - p_t)\delta_{js}(1 - \delta_{ms})\delta_{rt} + \mu^2 C_s C_t p_m - \mu C_t p_m (1 - p_s)\delta_{rs}$$

$$- \mu C_s p_m \delta_{jt}(1 - \delta_{mt}) + p_m(1 - p_s)\delta_{jt}\delta_{rs}(1 - \delta_{mt})]$$

$$\frac{p_j}{p_1}\left[\frac{p_r}{p_1}(1 - \delta_{r1}) + \delta_{mr}\right].$$

In breaking up the expression into a series of fourteen triple sums, and summing each up under restricted terms as defined by the δ functions, the whole expression simplifies to

$$\frac{\mu}{p_1}[C_s p_t^2 + C_t p_s^2], \qquad (3.8.5)$$

so that the asymptotic covariance term,

$$\operatorname{cov}\left(\frac{M_s}{\sqrt{L}}, \frac{M_t}{\sqrt{L}}\right) = \mu_{11}^{-1} B_{st}$$

$$= \mu^{(2)} C_s C_t - \mu(C_s^2 + C_t^2) + C_s p_t^2 + C_t p_s^2. \qquad (3.8.6)$$

Therefore, the asymptotic variance-covariance matrix, D^*, of $\underline{M}/\sqrt{L}$ is given by

$$d_{jj}^* = C_j[p_j^2 + (1 - p_j)^2] + C_j^2\left(\frac{\mu^{(2)}}{\mu} - 2\mu\right) \qquad (3.8.7)$$

$$d_{st}^* = C_s p_t^2 + C_t p_s^2 + \mu^{(2)} C_s C_t - \mu(C_s^2 + C_t^2) \qquad (3.8.8)$$

for $s \neq t$.

3.9 Variance Approximations

In using L_j/L to estimate F_j, recall from (3.4.4) that the asymptotic variance of L_j/L is

$$\frac{p_j(1-p_j)\left(1+\frac{\sigma^2}{\mu^2}\right)}{\left(\frac{L}{\mu}\right)} = \frac{p_j(1-p_j)\left(1+\frac{\sigma^2}{\mu^2}\right)}{E(N^*)} .$$

Under the assumption that the individual cells are not distinguishable, a value for N^* would be difficult to obtain. Since

$$\frac{M_j}{L} \xrightarrow{a.s.} \frac{p_j(1-p_j)}{\mu}$$

from (3.5.1), both

$$\text{(i)} \quad \sum_{j=1}^{k} \left(\frac{M_j}{1-\frac{L_j}{L}}\right)$$

and

$$\text{(ii)} \quad \frac{\sum_{j=1}^{k} M_j}{1-\sum_{j=1}^{k}\left(\frac{L_j}{L}\right)^2}$$

will be consistent estimators for $E(N^*)$, as would each individual estimator,

$$\frac{M_j}{\left(\frac{L_j}{L}\right)\left(1-\frac{L_j}{L}\right)}$$

for $j = 1,2,\ldots,k$. A comparison of the variances for each of these estimators could be obtained using the results of (3.8.7) and (3.8.8), but this becomes mathematically complex. It would appear that

$$\frac{\sum_{j=1}^{k} M_j}{1-\sum_{j=1}^{k}\left(\frac{L_j}{L}\right)^2}$$

would be the most stable of the available estimators, since it utilizes the sampling information from all k phases, and also $\sum_{j=1}^{k}\left(\frac{L_j}{L}\right)^2$ should stay fairly constant, even over widely ranging values for each L_j/L.

It follows that a reasonable consistent estimator of v_{jj}^* would be

$$\hat{v}_{jj}^* = \frac{\left(\frac{L_j}{L}\right)\left(1-\frac{L_j}{L}\right)\left(1+\frac{\sigma^2}{\mu^2}\right)}{\frac{\sum_{s=1}^{k} M_s}{1-\sum_{s=1}^{k}\left(\frac{L_s}{L}\right)^2}} \tag{3.9.1}$$

It still remains to find an estimate for σ^2/μ^2, or at least to get working approximations to its value.

A lower bound on the variance term would exist for a cell division resulting in a system of cells very close in shape and size, so that μ^2 would be considerably greater than σ^2, resulting in σ^2/μ^2 approaching zero.

One possible way to obtain estimates for μ and σ^2 is to use the information available from the different segments of different phases on the sampling transect. Let N_{nj} = the number of cell intercepts composing the n^{th} segment of phase j. Recall from before that N_{nj} has a geometric distribution with

$$E(N_{nj}) = \frac{1}{1 - p_j}$$

$$V(N_{nj}) = \frac{p_j}{(1 - p_j)^2} \quad .$$

Therefore, writing Z_{nj}, the length of the n^{th} segment of phase j, as

$$Z_{nj} = \sum_{i = a_{nj}+1}^{a_{nj} + N_{nj}} X_i \, ,$$

where a_{nj} = the number of cells preceding the n^{th} segment of phase j,

$$\mu_j = E(Z_{nj}) = E(N_{nj})\mu = \frac{\mu}{1 - p_j}$$

and

$$\sigma_j^2 = V(Z_{nj}) = V(N_{nj}\mu) + E(N_{nj}\sigma^2)$$

$$= \frac{p_j\mu^2}{(1-p_j)^2} + \frac{\sigma^2}{1-p_j} . \qquad (3.9.2)$$

Rewriting the first equation as

$$\mu = (1-p_j)E(Z_{nj}) ,$$

μ can be estimated from each color by using the average length of a phase j segment, $\bar{Y}_j$, multiplied by $(1 - \frac{L_j}{L})$. Thus a weighted average of estimates of μ over all colors would yield a consistent estimate for μ. Since $\sum_{j=1}^{k} \mu p_j = \mu$, one possible such estimate would be

$$\hat{\mu} = \sum_{j=1}^{k} \left(\frac{L_j}{L}\right)\left(1 - \frac{L_j}{L}\right)\bar{Y}_j . \qquad (3.9.3)$$

Rewriting (3.9.2) as

$$\sigma^2 = V(Z_{nj})(1-p_j) - \frac{p_j\mu^2}{1-p_j}$$

would yield an estimate of σ^2 by estimating p_j with L_j/L and μ by (3.9.3), and using the sample variance of the length of the phase j segments as an estimate for $V(Z_{nj})$.

3.10 The Results of Hilliard and Cahn

The method just suggested in the previous section is used by Hilliard and Cahn [8], who are responsible for many of the applied results in volume-fractions analysis. While their results are only derived for $k = 2$ phases, a similar argument could extend their results for $k > 2$ phases. To compare the earlier results with Hilliard and Cahn's results for $k = 2$, let $p_1 = p$, $p_2 = q = 1 - p$, so that

$$\mu_1 = \frac{\mu}{1-p}, \quad \mu_2 = \frac{\mu}{p},$$

$$\sigma_1^2 = \frac{\sigma^2}{1-p} + \frac{p\mu^2}{(1-p)^2},$$

and

$$\sigma_2^2 = \frac{\sigma^2}{p} + \frac{(1-p)\mu^2}{p^2}.$$

Therefore, by breaking down the asymptotic variance for $k = 2$ into an expression using the means and variances of the segments,

$$v_{jj}^* = \frac{p(1-p)\left[1 + \frac{\sigma^2}{\mu^2}\right]}{\frac{L}{\mu}}$$

$$= \frac{p^2(1-p)^2\left[1 + \frac{\sigma^2}{\mu^2}\right]}{\frac{L}{\frac{\mu}{1-p} + \frac{\mu}{p}}}.$$

$$= \frac{p^2(1-p)^2}{\frac{L}{\mu_1+\mu_2}} \left\{ \frac{\sigma^2+\mu^2}{\mu^2} \right\}$$

$$= \frac{p^2(1-p)^2}{\frac{L}{\mu_1+\mu_2}} \left[\frac{(1-p)\sigma^2+p\mu^2}{\mu^2} + \frac{p\sigma^2+(1-p)\mu^2}{\mu^2} \right]$$

$$= \frac{p^2(1-p)^2}{E(M_1)} \left[\frac{\sigma_1^2}{\mu_1^2} + \frac{\sigma_2^2}{\mu_2^2} \right], \qquad (3.10.1)$$

which is the Hilliard-Cahn result. The sample estimates gathered from the line segments would serve to estimate μ_1, μ_2, σ_1^2, and σ_2^2, while L_1/L and M_1 estimate p and $E(M_1)$ respectively. This approach for estimating the variance would probably yield accurate results for a random model with many shaded features, thus allowing for many segments of both colors on the line segment for estimating the segment parameters.

It should also be noted that the Hilliard-Cahn result can be obtained by looking at the segments on the line transect as an alternating renewal process. By defining

Y_i^1 = length of the i^{th} segment of phase 1

Y_i^2 = length of the i^{th} segment of phase 2

the Y_i^j's are i.i.d. with mean μ_j, and variance σ_j^2 for $j = 1,2$. With

$$\frac{L_1}{L} = \frac{\sum_{i=1}^{M_1} Y_i^1}{L} ,$$

a variance approximation given by Smith [23] yields

$$V\left(\frac{L_1}{L}\right) \doteq \frac{1}{L(\mu_1+\mu_2)}\left\{\sigma_1^2 - 2\sigma_1^2\frac{\mu_1}{\mu_1+\mu_2} + (\sigma_1^2+\sigma_2^2)\left(\frac{\mu_1}{\mu_1+\mu_2}\right)^2\right\}$$

$$= \frac{1}{L(\mu_1+\mu_2)}\left\{\sigma_1^2\left[1 - 2\frac{\mu_1}{\mu_1+\mu_2} + \left(\frac{\mu_1}{\mu_1+\mu_2}\right)^2\right] + \sigma_2^2\left(\frac{\mu_1}{\mu_1+\mu_2}\right)^2\right\}$$

$$= \frac{1}{L(\mu_1+\mu_2)}\left\{\sigma_1^2\left(\frac{\mu_2}{\mu_1+\mu_2}\right)^2 + \sigma_2^2\left(\frac{\mu_1}{\mu_1+\mu_2}\right)^2\right\}$$

$$= \frac{\mu_1^2\mu_2^2}{L(\mu_1+\mu_2)^3}\left\{\frac{\sigma_1^2}{\mu_1^2} + \frac{\sigma_2^2}{\mu_2^2}\right\}.$$

With

$$p = \frac{\mu_1}{\mu_1+\mu_2} ,$$

$$1-p = \frac{\mu_2}{\mu_1+\mu_2} ,$$

and

$$E(M_1) = \frac{L}{\mu_1+\mu_2} ,$$

the above reduces to

$$\frac{p^2(1-p^2)}{E(M_1)}\left[\frac{\sigma_1^2}{\mu_1^2}+\frac{\sigma_2^2}{\mu_2^2}\right],$$

which is (3.10.1).

For situations in which the Hilliard-Cahn procedure may yield innaccurate variance estimates due to a small number of segments on the transect, it may be helpful to study a possible model for the cell structure in order to get estimates for μ and σ^2. The next two sections present possible models for consideration in such a situation.

3.11 Poisson Lines Model

The mosaic that arises from a Poisson Process, known as the Poisson Lines Model, allows for the random division of space by a set of random lines, as detailed in Section 2.4. Letting τ be the density of the random lines on the process, a direct result of Miles [13] is that the intersection points of the line transect with the random lines forms a Poisson Process of density $\lambda = \frac{2\tau}{\pi}$. Therefore, the distribution of the X_i's will be exponential with parameter λ. With the exponential distribution, $\mu^2 = \sigma^2 = \frac{1}{\lambda^2}$, so that

$$v_{jj}^* = \frac{2p_j(1-p_j)}{E(N^*)}, \qquad (3.11.1)$$

where p_j can be estimated by L_j/L and $E(N^*)$ approximated, as mentioned before, by

$$\frac{\sum_{s=1}^{k} M_s}{1 - \sum_{s=1}^{k} \left(\frac{L_s}{L}\right)^2} ,$$

which generalizes Scheaffer's [20,21] results for $k = 2$. Numerical results will be presented in Chapter 7, which will allow for comparisons between this approximation and the Hilliard-Cahn technique from the previous section.

3.12 Meijering Cell Model

The Meijering Cell Model, as developed in Section 2.5, also results in a set of convex polygons with identical size distributions on the mosaic. The means and variances of the cell sizes are tabulated by Gilbert [6] for two and three-dimensional mosaics sampled by areal sections and lineal transects. For a two-dimensional model generated by a point process of intensity τ, sampled by a lineal transect, the intercept lengths of the line with the individual cells will have a mean $\mu = 1.027\tau^{-\frac{1}{2}}$ and variance $\sigma^2 = .188\tau^{-1}$. Therefore, for this model,

$$v_{jj}^* = \frac{1.18 p_j (1 - p_j)}{E(N^*)}$$

where $E(N^*)$ can be approximated as before.

3.13 Tests on the Mosaic Using Segment Counts

Consider the placing of multiple transects inside the boundaries of the mosaic. The measurements of interest include the individual segment counts for each phase from each transect. The asymptotic properties of the segments were developed in Section 3.8. This section will utilize the results from Section 3.8 to develop a test for the equality of mean segment counts, over all transects. Applications for a test of this kind would arise for feature orientation and phase homogeneity. The test for orientation would consist of independently placing n transects of length L at differing angles across the mosaic. If, for individual colors, the number of segments per unit length varies significantly from transect to transect, then a possible color orientation might exist.

A segregation test would involve the same ideas, except that the n transects would be placed in a parallel grid across the mosaic.

Let

$M_{j,m}$ = the number of segments of phase j on the m-th transect

and let

$$\underline{M}'_m = (M_{1,m}\ M_{2,m}\ \cdots\ M_{k,m}).$$

The general test, which follows Scheaffer's [19] test for k = 2, is set up as follows:

$$H_o: \quad E(\underline{M}_1) = E(\underline{M}_2) = \qquad = E(\underline{M}_n)$$

H_a: There is at least one j' such that $E(\underline{M}_{j'}) \neq E(\underline{M}_j)$ for $j \neq j'$.

Recall from Section 3.8, that if H_o is true, $\frac{1}{\sqrt{L}}(\underline{M}_m - L\underline{C}) \xrightarrow{D} N_k(\underline{0}, D^*)$ as $L \to \infty$, so that under H_o, $\frac{1}{L}(\underline{M}_m - L\underline{C})'(D^*)^{-1}(\underline{M}_m - L\underline{C})$ has asymptotically a central χ^2 distribution with k degrees of freedom. Let

$$\underline{\bar{M}} = \frac{1}{n} \sum_{m=1}^{n} \underline{M}_m$$

be the vector of averages for the number of segments for each phase over the n transects. Then $\frac{1}{\sqrt{L}}(\underline{\bar{M}} - L\underline{C})$ also has an asymptotic multivariate normal distribution with zero mean vector and variance-covariance matrix $\frac{1}{n} D^*$. Thus by writing

$$\underline{M}_m - L\underline{C} = (M_m - \underline{\bar{M}}) + (\underline{\bar{M}} - L\underline{C}),$$

it can be shown, using Cochran's Theorem, that under H_o,

$$T = \frac{1}{L} \sum_{m=1}^{n} (\underline{M}_m - \underline{\bar{M}})'(D^*)^{-1}(\underline{M}_m - \underline{\bar{M}})$$

has an asymptotic χ^2 distribution as $L \to \infty$ with $(n-1)k$ degrees of freedom. Therefore, the test statistic will be

$$T = \frac{1}{L} \sum_{m=1}^{n} (\underline{M}_m - \underline{\bar{M}})'(\hat{D}^*)^{-1}(\underline{M}_m - \underline{\bar{M}}),$$

where $\hat{D}^*$ must be a consistent estimator for D^*. Recall that

$$d^*_{jj} = C_j[p_j^2 + (1-p_j)^2] + C_j^2\left[\frac{\mu^{(2)}}{\mu} - 2\mu\right]$$

and

$$d^*_{st} = C_s p_t^2 + C_t p_s^2 + \mu^{(2)} C_s C_t - \mu(C_s^2 + C_t^2)$$

for $s \neq t$, so that p_j could be estimated by the total fraction of phase j over all n transects, and μ and $\mu^{(2)}$ estimated from the segment information, as developed in Section 3.9. Since M_j/L is consistent for C_j, the total number of phase j segments per unit length should serve as an estimate for C_j. With D^* approximated using the terms from above, the test for equal means would have a rejection region of the form

$$\hat{T} > \chi^2_{(n-1)k,\alpha}\,.$$

Recall from Section 3.9 and 3.10 the need for the assumption of random cell assignment in order to use the variance approximation from the Poisson Lines Model. An orientation test could be used as a prelude before such an approximation is used.

3.14 Spatial Covariance Function Approach to Using Lineal Sampling: Poisson Process

The approach discussed in Section 2.3 to obtain variance expressions for the areal (volume)-fraction estimators depends on obtaining an estimate for the covariance function for points on the mosaic. When estimates of the covariance function are available, then the corresponding variance for lineal fractions is

$$V\left(\frac{L_j}{L}\right) = \int_0^\infty \operatorname{cov}(h)b(h)dh ,$$

where b(h) represents the density function for the distance separating 2 points randomly located on the line transect. When the sampling scheme, S, is a line transect of length L, let Y_1 and Y_2 represent the distance of the two points from the beginning of the line. The distances Y_1 and Y_2 are independently distributed as uniform random variables on (0,L), so that for $0 \le h \le L$,

$$P(|Y_1 - Y_2| \le h) = P(Y_{(2)} - Y_{(1)} \le h)$$

$$= \int_0^L \int_{Y_2-h}^L 2!\frac{1}{L^2} dh = 1 - \frac{(L-h)^2}{L^2} .$$

Therefore

$$b(h) = \begin{cases} \frac{2}{L}(1 - \frac{h}{L}) & \text{for } 0 \le h \le L \\ 0 & \text{otherwise} . \end{cases} \qquad (3.14.1)$$

Now for any covariance function representing the underlying mosaic, the variance of L_j/L can be obtained.

The first model to be considered is the Poisson Lines Model, already having been defined in Section 2.4, for which the covariance function for color j is

$$\mathrm{cov}_j(h) = p_j(1 - p_j)e^{-\lambda h} ,$$

where λ is the underlying parameter of the process. Therefore,

$$V\left(\frac{L_j}{L}\right) = \frac{2p_j(1 - p_j)}{L} \int_0^L \left[e^{-\lambda h} - \frac{he^{-\lambda h}}{L}\right] dh ,$$

which as L gets large, will result in a comparable expression to (3.11.1).

3.15 Boolean Scheme

The model for the Boolean Scheme, as developed in detail in Section 2.6, yields a covariance function

$$\mathrm{cov}(h) = q^2 e^{\tau K(h)} - q^2$$

for a model generated from a point process of intensity τ. When S is a lineal transect of length L,

$$V\left(\frac{L_j}{L}\right) = \int_a [q_2 e^{\tau K(h)} - q^2] \frac{2}{L}(1 - \frac{h}{L}) dh .$$

One possible approach to integrating the above expression is to use a Taylor expansion, with

$$e^{\tau K(h)} = e^{\tau K(0)} + h\tau K'(0)e^{\tau K(0)} + o(h)$$

$$= \frac{1}{q}[1 + h\tau K'(0)] + o(h) ,$$

where K(h) is defined as in Section 2.6. Recall from Chapter 2 that $-\pi C'(0)$ equals the expected boundary length per unit area, which in turn can be estimated by $\frac{\pi}{2L} \times$ the number of intersections of the transect with the cell boundaries. With $-\pi C'(0) = -\pi q^2 \tau K'(0)e^{\tau K(0)} = -\pi q \tau K'(0)$, an estimate for $\tau K'(0)$ will be

$$\frac{\pi}{2L} M = \frac{-M}{2qL} .$$

With this approximation available, $e^{\tau K(h)}$ can be approximated by

$$\frac{1}{q}\left[1 - \frac{hM}{2qL}\right] , \qquad (3.15.1)$$

so that $V\left(\frac{L_j}{L}\right)$ can be approximated by

$$\int_0^L (q - \frac{hM}{2L} - q^2)\,\frac{2}{L}\,(1 - \frac{h}{L})dh$$

$$= \int_0^L pq\,\frac{2}{L}\,(1 - \frac{h}{L}) - \frac{hM}{L^2}(1 - \frac{h}{L})dh$$

$$= pq - \frac{M}{2} + \frac{M}{3} = pq - \frac{M}{6} \qquad (3.15.2)$$

However, a problem with this approximation might arise if $K(h) = 0$ when h is greater than some fixed distance.

A second possibility would involve studying specific models which could generate a mosaic from the definition of the Boolean scheme. Unfortunately, many modeling assumptions result in an expression for K(h) that makes $V\left(\frac{L_j}{L}\right)$ mathematically complex. One of the simpler models to work with assumes each β_j set to be a ball of radius r. With $K(0) = \pi r^2$, it can be shown using geometric results mentioned by Matern [11], that

$$K(h) = \begin{cases} K(0) - 2r^2 \sin^{-1}\frac{h}{2r} - rh\sqrt{1 - \frac{h^2}{4r^2}} & \text{for } 0 \le h \le 2r \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$V\left(\frac{L_j}{L}\right) \doteq \int_0^L [q^2 e^{\tau K(h)} - q^2]\frac{2}{L}(1 - \frac{h}{L})dh$$

$$= \int_0^{2r} \left(qe^{\tau\left[-2r^2\sin^{-1}\frac{h}{2r} - rh\sqrt{1 - \frac{h^2}{4r^2}}\right]} - q^2\right) \frac{2}{L}(1 - \frac{h}{L})dh,$$

which is mathematically difficult to integrate. An approximation can be obtained through a Taylor expansion of the term

$$e^{\tau\left[-2r^2\sin^{-1}\frac{h}{2r} - rh\sqrt{1 - \frac{h^2}{4r^2}}\right]}.$$

However, using the approximation from (3.15.1), $V\left(\frac{L_j}{L}\right)$ can be approximated by

$$\int_0^{2r} pq \frac{2}{L}(1 - \frac{h}{L}) - \frac{hM}{L^2}(1 - \frac{h}{L})dh$$

$$= \frac{pq(4r)}{L} - \frac{pq8r^2}{L^2} - \frac{2r^2M}{L^2} + \frac{8Mr^3}{3L^3} .$$

As L gets large, the second and fourth terms should get very small, so that $V\left(\frac{L_j}{L}\right)$ can be approximated by

$$\frac{pq(4r)}{L} - \frac{2r^2M}{L^2} . \qquad (3.15.4)$$

Note that M is a function of L, so that the second term in the expression is really the same order of L as in the first term, so that as L gets large, the two terms should get close. However, it should be realized that the whole approximation is based on the expansion of $e^{\tau k(h)}$, from which estimates are only obtained for the first two terms. If any additional terms of the expansion are significantly different from zero, the corresponding variance estimate may not yield an accurate approximation.

CHAPTER 4

AREAL SAMPLING

4.1 Introduction

In considering the parameters of the n-dimensional mosaic as defined in Section 3.1, it was mentioned that several different sampling procedures would be considered for obtaining information vital to the estimation of various properties of the mosaic. Chapter 3 developed the ideas behind the lineal sampling approach, with the asymptotic properties of the lineal estimators being developed from a renewal theory argument and the use of some limit theorems on Markov Renewal Processes. This chapter will develop estimators for the mosaic parameters when the sampling procedure involves the random location of a planar section or sections of fixed total area A entirely within the borders of the mosaic. The fraction of total area of the planar features of phase j gives an unbiased and consistent estimate of the fraction of the j^{th} phase on the entire mosaic. As done in Chapter 3, the next step will be the development of the asymptotic joint distribution for the vector of areal-fraction estimators, followed by estimation of the variance components through the utilization of other data obtainable from the areal sample.

4.2 Asymptotic Distribution of $\underline{A}/\sqrt{A}$

Many of the arguments used in Chapter 3 can be extended and/or paralleled to derive the asymptotic distribution of the vector of areal-fraction estimators, assuming that the sample is a planar section of area A randomly selected from the mosaic. Under the assumption of the underlying cell model for the construction of the mosaic, as developed in Section 3.2, let

X_i = area on the i^{th} cell on the plane,

$$E(X_i) = \mu < \infty \ , \qquad V(X_i) = \sigma^2 < \infty \ ,$$

$$\phi_{ij} = \begin{cases} 1 & \text{if the } i^{th} \text{ cell is phase } j \\ 0 & \text{otherwise ,} \end{cases}$$

N^* = the number of cells on the sample planar section .

Again, assuming that the individual cells will not generally be distinguishable, N^* will not be available from the data. Letting

$$A_j = \sum_{i=1}^{N^*} \phi_{ij} X_i \ ,$$

A_j will represent the total sampled area due to phase j features and A_j/A will be a consistent estimator for p_j, for $j = 1,2,\ldots,k$. Let $\underline{A}' = (A_1 \ A_2 \ \cdots \ A_k)$. The approach used to develop the asymptotic distribution of $\frac{1}{\sqrt{A}} \underline{A}$ is to think of the cells being dropped sequentially and filling up a total area A, much as the line segments fill up the lineal transect of length L. Assuming that the individual cells will be small in comparison to the large sampling area, so that of the sample will consist a large number of cells, any border effect of the areal sample will be negligible. With the independent assignment of colors to the cells, the sequential arrangement of cells will allow for parallel arguments to the renewal theory derivations of Chapter 3. In addition to A_j/A being a consistent estimator for p_j, it also is asymptotically unbiased, this being a direct result of Wald's Theorem in an analagous proof to the lineal case of Section 3.3.

Furthermore, the joint asymptotic distribution of $\frac{1}{\sqrt{A}}(\underline{A} - \underline{p})$ will be normal with zero mean vector and variance-covariance matrix V_A, so that the variance of A_j/A can be approximated for large A by

$$v_{jj}^{**} = \frac{p_j(1 - p_j)\left(1 + \frac{\sigma^2}{\mu^2}\right)}{A/\mu} \qquad (4.2.1)$$

and covariance term,

$$v_{st}^{**} = \frac{-p_s p_t\left(1 + \frac{\sigma^2}{\mu^2}\right)}{A/\mu} \qquad (4.2.2)$$

for $s \neq t$. Hence the next step is to get practical estimates for μ and σ^2 from the information available from the sample data.

4.3 Estimation of the Variance Components

To obtain a practical estimate of the terms of V^{**}, A_j/A and $1 - A_j/A$ serve as consistent estimators for p_j and $1 - p_j$ respectively, but the other terms are parameters of the underlying scheme.

Some models may arise in which the cells are very similar in both size and shape. In this case, $E(X_i^2)/\mu^2$ is very close to one, so a lower bound on v_{jj}^{**} is

$$\frac{\left(\frac{A_j}{A}\right)\left(1 - \frac{A_j}{A}\right)}{A/\mu} .$$

Since $E(N^*)/A \to 1/\mu$ as $A \to \infty$, A/μ can be thought of as being approximately equal to $E(N^*)$, and an estimate for $E(N^*)$ can be obtained, similar to the one from Section 3.9. Defining

M_j = the number of phase j features on the sample,

$$\sum_{j=1}^{k} M_j \bigg/ \left(1 - \sum_{j=1}^{k} \left(\frac{A_j}{A}\right)^2\right)$$

is a consistent estimator of $E(N^*)$.

However, the feature count in two dimensions is not as easily obtained as the intercept count in one dimension since very irregularly shaped phases make counting difficult. Frequently, the planar sample section may have only one continuous feature for a certain phase, such as in viewing a large body of water in a land study situation.

Perhaps a more useful result could be obtained from a basic stereological result introduced in Chapter 2. Recall that E[boundary length for phase j features]/A = $-\pi C_j'(0)$, where $C_j(h)$ is as defined in Section 2.3, for phase j. For this estimate to be used, an underlying process for cell modeling must be assumed, so that a form for $C_j(h)$ can be obtained. The next section discusses this estimate for a frequently used model, the Poisson Lines Model.

4.4 Poisson Process

If the underlying process constitutes the realization of a Poisson Process of density $\lambda^* = \tau/\pi$, as stated by Miles [13], several additional results of Miles will help to simplify the asymptotic variance expressions.

The individual cell areas have mean

$$\mu = \frac{1}{\pi(\lambda^*)^2}$$

and variance

$$\sigma^2 = \frac{\pi^2 - 2}{2\pi^2(\lambda^*)^4}$$

and the expected cell boundary length is $2/\lambda^*$. With $\sigma^2/\mu^2 = (\pi^2 - 2)/2$,

$$v_{jj}^{**} = \frac{\pi^2 p_j(1 - p_j)}{2E(N^*)} = \frac{\pi p_j(1 - p_j)}{2A(\lambda^*)^2} . \qquad (4.4.1)$$

Recall from Section 3.14 that for a Poisson Lines Model generated from a point process of intensity τ, the covariance function of color j is

$$\mathrm{cov}_j(h) = p_j(1 - p_j)e^{-\frac{2\tau}{\pi}h} = p_j(1 - p_j)e^{-2\lambda^* h} ,$$

so that the expected boundary length per unit area of phase j features is

$$\frac{E(B_j)}{A} = 2\lambda^* p_j(1 - p_j) . \qquad (4.4.2)$$

Therefore,

$$\frac{B_j}{2A\pi p_j(1 - p_j)} \qquad (4.4.3)$$

gives an unbiased estimate of λ^*.. Utilizing the sampling information over all colors,

$$\frac{1}{2Ak\pi} \frac{\sum_{j=1}^{k} B_j}{1 - \sum_{j=1}^{k} \left(\frac{A_j}{A}\right)^2}$$

gives a consistent estimate for λ^*. In arguments similar to those of Section 3.9, it would seem that this would give a reasonable estimate of λ, since the denominator expression of $\sum_{j=1}^{k} \left(\frac{A_j}{A}\right)^2$ would appear to be fairly stable over varying areal-fraction estimates.

4.5 Meijering Cell Model

For the Meijering Cell Model defined in Section 2.5, Gilbert [6] tabled the following mean and variance for the areas of cells arising from a point process of intensity τ:

$$\mu = 1.018\tau^{-\frac{2}{3}}$$

and

$$\sigma^2 = .228\tau^{-\frac{4}{3}}.$$

Therefore, for this model

$$v_{jj}^{**} = 1.22 \frac{p_j(1-p_j)}{E(N^*)} \quad .$$

It should be noted that for both the lineal and areal cases, this model yields less variability for the areal-fraction estimators than the Poisson Lines Model, so that it might be preferred for mosaics with similar type features.

4.6 Comparisons with the Results of Hilliard and Cahn

A variance expression for A_j/A is obtained by Hilliard and Cahn under the assumption of two rigid conditions. The first assumption is that the number of equal sized features of color j within the given region follows a Poisson distribution. Second, the areal fraction for any phase is assumed to be very small. Their final result is

$$V\left(\frac{A_j}{A}\right) = \frac{E(M_j)}{A^2}[\sigma_j^2 + \mu_j^2] \qquad (4.5.1)$$

where μ_j and σ_j^2 denote respectively the mean and variance of a phase j feature. Recalling that a feature may be composed of one or more individual cells, and using the arguments of Section 3.9, it can be shown that

$$\mu_j = \frac{\mu}{1 - p_j}$$

and

$$\sigma_j^2 = \frac{\sigma^2}{1 - p_j} + \frac{p_j \mu^2}{(1 - p_j)^2} ,$$

or equivalently,

$$\mu = \mu_j (1 - p_j)$$

and

$$\sigma^2 = (1 - p_j)[\sigma_j^2 - p_j \mu_j^2],$$

with

$$v_{jj}^{**} = \frac{p_j (1 - p_j)}{A/\mu} \left[1 + \frac{\sigma^2}{\mu^2} \right]$$

$$= \frac{A}{\mu} p_j (1 - p_j) \frac{\mu^2}{A^2} \left[1 + \frac{\sigma^2}{\mu^2} \right]$$

$$= \frac{E(M_j)}{A^2} [\mu^2 + \sigma^2]$$

$$= \frac{E(M_j)}{A^2} [\sigma_j^2 + \mu_j^2 + \{(-3p_j + 2p_j^2)\mu_j^2 - p_j \sigma_j^2\}] \tag{4.6.2}$$

which reduces to (4.6.1) when p_j is very small. As a result, the variance expression obtained by Hilliard and Cahn represents a special case of an asymptotic variance from a more generalized model. The result of (4.6.1) may be a useful expression if there are a few individual features from which we can obtain areal measurements within the sampling plane. The sample mean and variance can then be used as consistent estimators for μ_j and σ_j^2 respectively.

4.7 Spatial Covariance Function Approach

Recall from Section 2.2 that if S represents the sampling scheme on the mosaic, then the variance of the areal-fraction estimate is

$$\int_G \operatorname{cov}(h) b_S(h) ,$$

where $b_S(h)$ is the density function for h, the distance between 2 points randomly located on S. Even for a recognizable covariance function, the variance term may be extremely complicated mathematically. The one modelling scheme that has been studied in the results of Chapter 3 and 4 has been the Poisson Lines Model, in which for color j,

$$\operatorname{cov}_j(h) = p_j(1 - p_j)e^{-2\lambda^* h} .$$

In studying the areal-sampling approach to volume fraction estimation, let S represent a convex set of total area A on the plane. A geometrical result due to Borel and given by Matérn [11] is that

$$b_S(h) = \frac{2\pi h}{A} - \frac{2C_S h^2}{A^2} + o(h^2) \qquad (4.7.1)$$

as h → 0, where

C_S = total perimeter of boundary length of S.

Note that this approximation is only applicable when h is small, but as h gets large, $\text{cov}_j(h) = p_j(1 - p_j)e^{-2\lambda^* h}$ gets very small. Also, as A gets large, the second term of $b_S(h)$ becomes very small, so that $V\left(\frac{A_j}{A}\right)$ can be approximated by

$$\frac{2\pi p_j(1 - p_j)}{A} \int^{\infty} h e^{-2\lambda^* h} dh$$

$$= \frac{p_j(1 - p_j)}{2(\lambda^*)^2 A} ,$$

yielding the same result as (4.4.1).

CHAPTER 5

POINT SAMPLING

5.1 Introduction

The final method of estimation to be discussed involves the use of point sampling on the mosaic. By placing points within the boundaries of the mosaic, the fraction of points falling upon a certain phase can be used as an estimator of the phase proportion. Koop [9] has discussed 21 methods of point sampling for the two-phase planar mosaic. Several of the simpler techniques will be discussed and extended for the general k-phased mosaic. Also, the asymptotic distribution for the phase proportion estimators will be developed under restricted conditions.

5.2 Random Points on the Plane

For a planar mosaic of area A, let $A_1, A_2, \ldots, A_k$ denote the total area on the mosaic for phases 1,2,...,k respectively, such that $\sum_{j=1}^{k} A_j = A$. As noted earlier in Chapter 3, the mosaic itself may represent a planar section of a two or three-dimensional population figure, for which each A_j/A would serve as an estimator for the population proportion of phase j. For a point located at random within the boundaries of the planar mosaic, the probability that the point falls on a phase j feature is $A_j/A = F_j$. Random sampling involves the independent location of a fixed number, n, of points at random. Let N_j equal the number of the n points belonging to phase j features for $j = 1,2,...,k$. The distribution of $\underline{N}' = (N_1 \; N_2 \; \cdots \; N_k)$ is multinomial with probabilities $F_1,...,F_k$. This is, of course, a standard result of random point sampling. Individually, N_j/n is an unbiased estimator of F_j, with

$$V\left(\frac{N_j}{n}\right) = \frac{F_j(1-F_j)}{n} \tag{5.2.1}$$

and

$$\operatorname{cov}\left(\frac{N_s}{n}, \frac{N_t}{n}\right) = \frac{-F_sF_t}{n} \tag{5.2.2}$$

for $s \neq t$. As n gets large, the distribution of $\underline{N}/\sqrt{n}$ approaches normality. For random points sampled from a lineal transect with phase proportions $\frac{L_1}{L},...,\frac{L_k}{L}$ respectively, the results will be similar, except that L_j/L will be used instead of F_j.

5.3 Systematic Points on a Line

A second method of point sampling that is widely used in many different kinds of experiments involves systematic sampling of points on one or more lines passing through the mosaic. This method has some practical advantages in its use over the random sampling technique. It generally will offer the experimenter a more convenient design or layout of points on the mosaic, and estimators obtained from systematic point counts also are unbiased, with variances inversely proportional to n.

For an individual line, denote the lineal fraction of phase j by L_j/L, as defined in Chapter 3. It should be mentioned that two possible situations may exist to allow for systematic sampling. The first will occur if the line by itself constitutes a one-dimensional mosaic, in which the vector $\underline{L}$ is constant. The other situation arises when the transect is randomly located in a higher-dimensioned mosaic, as discussed in Chapter 3. This will be the case under consideration in this section, so that for each transect, the values for $\{L_j\}_{j=1}^{k}$ will be random, with L_j/L representing only an estimator of F_j. The process of locating a fixed number, n, of points on the line will start by dividing the line into n intervals, each interval of width $\Delta = L/n$. The first point is then placed at random in the first interval, with each succeeding point a distance Δ from its nearest neighbors, as defined by Cox [3].

Define

$$u_{ij} = \begin{cases} 1 & \text{if the } i^{th} \text{ point on the line belongs to a phase } j \text{ segment} \\ 0 & \text{otherwise.} \end{cases}$$

Due to the random location of the line and the first point, $E(u_{ij}) = F_j$ and $V(u_{ij}) = F_j(1 - F_j)$. Letting

$$\frac{N_j}{n} = \frac{\sum_{i=1}^{n} u_{ij}}{n} ,$$

$E\left(\frac{N_j}{n}\right) = F_j$, so that N_j/n would appear to be a reasonable estimator of F_j, since it is unbiased. To look at $V\left(\frac{N_j}{n}\right)$, it is first necessary to develop the

$$\mathrm{cov}(u_{ij}\ ,\ u_{i+s,j}) = E(u_{ij}, u_{i+s,j}) - F_j^2\ .$$

Define

$$\pi(h) = P(\text{2 points a distance } h \text{ apart belong to the same cell}).$$

Now,

$$\begin{aligned} E(u_{ij}, u_{i+s,j}) &= P(u_{ij} = 1\ ,\ u_{i+s,j} = 1) \\ &= P(u_{ij} = 1)P(u_{i+s,j} = 1 \mid u_{ij} = 1) \\ &= F_j[\pi(s\Delta) + \{1 - \pi(s\Delta)\}F_j] \\ &= F_j(1 - F_j)\pi(s\Delta) + p_j^2\ . \end{aligned}$$

Therefore,

$$\mathrm{cov}(u_{ij}\ ,\ u_{i+s,j}) = F_j(1 - F_j)\pi(s\Delta) \qquad (5.3.1)$$

so that,

$$\mathrm{Var}\left(\frac{N_j}{n}\right) = \frac{F_j(1 - F_j)}{n} + \frac{2}{n^2}\sum_{m=1}^{n} F_j(1-F_j)(n-m)\pi(m\Delta). \qquad (5.3.2)$$

To make use of the usual inferential techniques, an estimate for $\pi(m\Delta)$ must be found. The approach is intuitively parallel to the derivation of Section 3.9.

5.4 Estimation of the Variance

Define

$$\phi^j_{st} = \begin{cases} 1 & \text{if } u_{sj} = 1 \text{ and } u_{tj} = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$M^j_s = \sum_{i=1}^{n-s} \phi^j_{i,i+s}$$

can be used to study the transition out of phase j for points $s\Delta$ apart.

$$\begin{aligned} E(M^j_s) &= (n-s)E(\phi^j_{i,i+s}) \\ &= (n-s)P(u_{sj} = 1)P(u_{tj} = 0 \mid u_{sj} = 1) \\ &= (n-s)F_j[1-\pi(s\Delta)](1-F_j) \\ &= (n-s)F_j(1-F_j)[1-\pi(s\Delta)] \qquad (5.4.1) \end{aligned}$$

so that

$$\pi(s\Delta) = 1 - \frac{E(M^j_s)}{(n-s)F_j(1-F_j)} . \qquad (5.4.2)$$

Therefore, $V\left(\frac{N_j}{n}\right)$ can be estimated by

$$\frac{\left(\frac{N_j}{n}\right)\left(1-\frac{N_j}{n}\right)}{n} + \frac{2\left(\frac{N_j}{n}\right)\left(1-\frac{N_j}{n}\right)}{n^2} \sum_{m=1}^{n} (n-m)\left[1 - \frac{M^j_s}{(n-m)\left(\frac{N_j}{n}\right)\left(1-\frac{N_j}{n}\right)}\right]$$

$$= \frac{N_j}{n}\left(1 - \frac{N_j}{n}\right) - \frac{2}{n^2}\sum_{s=1}^{n-1} M_s^j \,. \qquad (5.4.3)$$

Note that as n gets large, Δ becomes smaller, so that the points get closer and closer on the line, allowing for more transitions in and out of a particular phase. However, in looking at (5.4.3), each variance approximation uses an estimate of $\pi(s\Delta)$ in terms of the individual phase j. Rather than use different estimates in each individual variance expression, it would seem far more reasonable to utilize all of the information in some form of weighted expression that could be used for all variance and covariance terms, independent of the particular phase under analysis. Following the same line of reasoning employed in the lineal case,

$$(n - s)\pi(s\Delta)F_j(1 - F_j) = (n - s)F_j(1 - F_j) - E(M_s^j)$$

for $j = 1,2,\ldots,k$, so that

$$(n - s)\pi(s\Delta)\left[1 - \sum_{j=1}^{k} F_j^2\right] = (n - s)\left[1 - \sum_{j=1}^{k} F_j^2\right] - \sum_{j=1}^{k} E(M_s^k)$$

resulting in

$$\hat{\pi}(s\Delta) = 1 - \frac{\sum_{t=1}^{k} M_s^t}{(n - s)\left[1 - \sum_{t=1}^{k}\left(\frac{N_t}{n}\right)^2\right]} \qquad (5.4.4)$$

as a stable estimate for $\pi(s\Delta)$. Averaging over all values of s, $V\left(\frac{N_j}{n}\right)$ can now be estimated by

$$\frac{\left(\frac{N_j}{n}\right)\left(1-\frac{N_j}{n}\right)}{n} + \frac{2}{n^2}\left(\frac{N_j}{n}\right)\left(1-\frac{N_j}{n}\right)\sum_{s=1}^{n-1}\left\{1-\frac{\sum_{t=1}^{k} M_s^t}{1-\sum_{t=1}^{k}\left(\frac{N_t}{n}\right)^2}\right\} \qquad (5.4.5)$$

5.5 Covariance Between Colors

If comparisons are to be made between colors, the covariance term needs to be established. For $s \neq t$,

$$E\left(\frac{N_s}{n}\,\frac{N_t}{n}\right) = \frac{1}{n^2}E\left[\left(\sum_{i=1}^{n} u_{is}\right)\left(\sum_{i=1}^{n} u_{it}\right)\right]$$

$$= \frac{1}{n^2}E\left(\sum_{i=1}^{n} u_{is}\,u_{it} + \sum_{i\neq m} u_{is}\,u_{mt}\right).$$

The first term of the above expression is zero, since a point cannot be in two different colors. Thus,

$$E\left(\frac{N_s}{n}\,\frac{N_t}{n}\right) = \frac{1}{n^2}\sum_{i\neq m} P(u_{is}=1\,,\,u_{mt}=1)$$

$$= \frac{2}{n^2}\sum_{m=1}^{n-1} F_s[1-\pi(m\Delta)]F_t(n-m)$$

$$= \frac{2}{n^2}F_sF_t\left[\sum_{m=1}^{n-1}(n-m) - \sum_{m=1}^{n-1}(n-m)\pi(m\Delta)\right]$$

$$= \frac{n-1}{n}F_sF_t - \frac{2}{n^2}F_sF_t\sum_{m=1}^{n-1}(n-m)\pi(m\Delta)\,. \qquad (5.5.1)$$

Therefore,

$$\operatorname{cov}\left(\frac{N_s}{n}, \frac{N_t}{n}\right) = \frac{-F_sF_t}{n} - \frac{2}{n^2} F_sF_t \sum_{m=1}^{n-1} (n - m)\pi(m\Delta) \qquad (5.5.2)$$

with estimation of $\pi(m\Delta)$ from (5.4.4).

5.6 Asymptotic Distribution of the Estimates

It will be shown that the distribution of $\frac{1}{\sqrt{n}}\underline{N}$ will be asymptotically normal under the assumption of the development of the mosaic from a Poisson Lines Model. Unlike the other areal (volume)-fraction estimators that have been obtained through lineal, areal, and random point measurements the systematic point count estimators cannot be written as an average of independent indicator functions. As discussed in Sections 5.3 and 5.4, the systematic points on the line are dependent with covariance function $\operatorname{cov}_j(h) = F_j(1 - F_j)\pi(h)$ for points h units apart. Therefore, the asymptotic normality of the vector of estimates must be achieved through a Central Limit Theorem for dependent random variables.

The theorem to be used is a multivariate result of Hannan [7], stated as follows:

Let $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$ be k × 1 vectors, defined by

$$\underline{X}_i = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ik} \end{pmatrix} = \begin{pmatrix} u_{i1} - F_1 \\ u_{i2} - F_2 \\ \vdots \\ u_{ik} - F_k \end{pmatrix},$$

such that

(i) Each vector of observations satisfies the strong mixing property.

(ii) $$\mu_{rsw}(m_1,m_2) = E(X_{1r}\ X_{m_1 s}\ X_{m_2 w})$$

for m_1 and m_2 integers such that $1 \le m_1 \le m_2 \le n$ is finite when summed over all possible values for m_1 and m_2.

(iii) The covariance spectrum, $[f(\varepsilon)]$ is continuous at $\varepsilon = 0$ and trace $[f(\varepsilon)]$ is uniformly bounded, where

$$f_{rs}(\varepsilon)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ih\varepsilon} C_{rs}(h)dh . \qquad (5.6.1)$$

Then, the vector

$$\sqrt{n}\begin{pmatrix}\bar{X}_1\\ \bar{X}_2\\ \vdots\\ \bar{X}_k\end{pmatrix} = \sqrt{n}\begin{pmatrix}\frac{N_1}{n} - F_1\\ \frac{N_2}{n} - F_2\\ \vdots\\ \frac{N_k}{n} - F_k\end{pmatrix}$$

will be asymptotically normal as $n \to \infty$ with variance-covariance matrix $2\pi[f(0)]$.

In the general case, the first condition (i) will not necessarily be true and can be extremely tedious to prove from a mathematical approach. However, under the assumption of a Poisson Lines Model used for the development of the cell structure of the mosaic (Section 2.4), the number of intersections the transect makes with the cell boundaries has a Poisson distribution. With this additional property, the combination of two theorems of Switzer [24] and Billingsley [2] yield the strong mixing property. Switzer proved that for the ordered set of collinear points on a planar Poisson Process, $\{u_{ij}\}_{i=1}^{n}$ will have the Markov property for $j = 1, 2, \ldots, k$, and thus the set of $\underline{X}_i$ vectors will satisfy the Markov property. The general strong mixing property will be satisfied if for two events, $\mathfrak{A}$ defined as a function of the information from the first i points, and $\mathfrak{B}$ defined as a function of the information from the points after the $(i+m-1)^{th}$ point, then

$$|P(\mathfrak{A}\mathfrak{B}) - P(\mathfrak{A})P(\mathfrak{B})| \leq \alpha(m) \downarrow 0,$$

as $m \to \infty$. Billingsley [2] showed that a stationary Markov Process satisfies the strong mixing property, so that condition (i) has been satisfied for the Poisson Lines Model. For this model, the covariance function is

$$\text{cov}_j(h) = F_j(1 - F_j)e^{-\lambda h}$$

which will now be used in showing that conditions (ii) and (iii) both hold.

To show the existence of the second condition, consider

$$I_{rsw}(m_1, m_2) = E(u_{1r}\, u_{m_1 s}\, u_{m_2 w}) \tag{5.6.2}$$

and

$$I_{rs}(m_i,m_j) = E(u_{m_i r}\, u_{m_j s}) . \qquad (5.6.3)$$

Then

$$\mu_{rsw}(m_1,m_2) = I_{rsw}(m_1,m_2) - F_r I_{sw}(m_1,m_2)$$

$$- F_s I_{rw}(1,m_2) - F_w I_{rs}(1,m_1) + 2F_r F_s F_w . \qquad (5.6.2)$$

Note that r,s,w, represent different phases on the mosaic so that several cases must be studied.

CASE I: Consider $r = s = w$. Without loss of generality, call this phase 1. Then

$$I_{11}(1,m_1,m_2) = P(u_{11} = 1 , u_{m_1 1} = 1 , u_{m_2 1} = 1),$$

which can occur in any one of three ways:

A: All three points belong to the same phase 1 cell.

B: The first two points belong to the same phase 1 cell, while the third belongs to a different phase 1 cell.

C: The second and third points fall in the same phase 1 cell, but the first point is on a different phase 1 cell.

D: All three points are on different phase 1 cells.

Therefore,

$$I_{111}(1,m_1,m_2) = P(A) + P(B) + P(C) + P(D)$$

where

$$P(A) = F_1 e^{-\lambda(m_2-1)\Delta},$$

$$P(B) = F_1^2 e^{-\lambda(m_1-1)\Delta}\left[1 - e^{-\lambda(m_2-m_1)\Delta}\right],$$

$$P(C) = F_1^2\left[1 - e^{-\lambda(m_1-1)\Delta}\right] e^{-\lambda(m_2-m_1)\Delta}$$

and

$$P(D) = F_1^3\left[1 - e^{-\lambda(m_1-1)\Delta}\right]\left[1 - e^{-\lambda(m_2-m_1)\Delta}\right].$$

Therefore,

$$I_{111}(m_1,m_2) = e^{-\lambda(m_1-1)\Delta}(F_1^2 - F_1^3) + e^{-\lambda(m_2-1)\Delta}$$
$$(F_1 - F_1^2 - F_1^2 + F_1^3) + e^{-\lambda(m_2-m_1)\Delta}(F_1^2 - F_1^3) + F_1^3, \quad (5.6.5)$$

$$I_{11}(m_1,m_2) = F_1 e^{-\lambda(m_2-m_1)\Delta} + F_1^2\left[1 - e^{-\lambda(m_2-m_1)\Delta}\right]$$
$$= F_1(1 - F_1)e^{-\lambda(m_2-m_1)\Delta} + F_1^2 . \qquad (5.6.6)$$

When these expressions are substituted into the expression for $\mu_{111}(m_1,m_2)$,

$$\mu_{111}(m_1,m_2) = G_1(F_1)e^{-\lambda\Delta(m_2-m_1)}$$

where $G_1(F_1)$ is a third degree polynomial function of F. To prove that this moment is finite over all values of m_1 and m_2, observe that

$$\sum_{m_2=1}^{\infty}\sum_{m_1=1}^{m_2} G_1(F_1)e^{-\lambda\Delta(m_2-1)} = \sum_{m_2=1}^{\infty} m_1 G(F_1)e^{-\lambda\Lambda(m_2-1)} < \infty$$

by result of the ratio test.

CASE II: Let $r = s \neq w$. To simplify notation, let $r = 1$, $w = 2$,

$$I_{112}(m_1,m_2) = P(u_{11} = 1 , u_{m_1,1} = 1 , u_{m_2,2} = 1),$$

which occurs when

A: The first two points belong to the same cell of phase 1, and the third in a phase 2 cell.

B: All three points fall in different cells, the first two being phase 1 cells and the third being a phase 2 cell.

$$P(A) = F_1 e^{-\lambda(m_1-1)\Delta}\left[1 - e^{-\lambda(m_2-m_1)\Delta}\right]F_2 , \tag{5.6.7}$$

$$P(B) = F_1\left[1 - e^{-\lambda(m_1-1)\Delta}\right]F_1\left[1 - e^{-\lambda(m_2-m_1)\Delta}\right]F_2 . \tag{5.6.8}$$

Therefore, $\mu_{112}(m_1,m_2)$ again reduces to the form of

$$G_2(F_1,F_2)e^{-\lambda(m_2-1)\Delta} ,$$

which when summed over all values for m_1 and m_2 is finite.

CASE III: Let $s = w \neq r$, and let $r = 1$ and $s = 2$. Then

$$I_{122}(m_1,m_2) = P(u_{11} = 1 , u_{m_1,2} = 1 , u_{m_2,2} = 1),$$

which occurs when:

A: The second and third points share a common phase 2 cell.

B: All three points appear in different cells.

$$P(A) = F_1\left[1 - e^{-\lambda(m_1-1)\Delta}\right]F_2 e^{-\lambda(m_2-m_1)\Delta} . \qquad (5.6.9)$$

$$P(B) = F_1\left[1 - e^{-\lambda(m_1-1)\Delta}\right]F_2\left[1 - e^{-\lambda(m_2-m_1)\Delta}\right]F_2 . \qquad (5.6.10)$$

Therefore,

$$I_{22}(m_1,m_2) = -F_1F_2^2e^{-\lambda(m_1-1)\Delta} - F_1F_2(1 - F_2)e^{-\lambda(m_2-1)\Delta}$$

$$+ F_1F_2(1 - F_2)e^{-\lambda(m_2-m_1)\Delta} ,$$

so that

$$\mu_{122}(m_1,m_2) = G_3(F_1,F_2)e^{-\lambda(m_2-1)\Delta} ,$$

and

$$\sum_{m_2=1}^{\infty} \sum_{m_1=1}^{m_2} \mu_{122}(m_1,m_2) < \infty .$$

CASE IV: Let $r \neq s \neq w$. To simplify notation, let $r = 1$, $s = 2$, $w = 3$,

$$I_{123}(m_1,m_2) = P(u_{11} = 1 , u_{m_1,2} = 1 , u_{m_2,3} = 1)$$

$$= F_1F_2F_3\left[1 - e^{-\lambda(m_1-1)\Delta}\right]\left[1 - e^{-\lambda(m_2-m_1)\Delta}\right] . \qquad (5.6.11)$$

Therefore,

$$\mu_{123}(m_1,m_2) = G_4(F_1,F_2,F_3)e^{-\lambda(m_2-1)\Delta} ,$$

which is finite when summed over m_1 and m_2, and the condition (ii) has been satisfied.

The third condition needed to satisfy Hannan's theorem states that the matrix of spectral density functions, $[f(\varepsilon)]$, is continuous at $\varepsilon = 0$ and that the trace $[f(\varepsilon)]$ is uniformly bounded over all values of ε. As previously defined by (5.6.1),

$$f_{rr}(\varepsilon) = \frac{F_r(1 - F_r)}{2\pi} \int_{-\infty}^{\infty} e^{-ih\varepsilon} e^{-\lambda\Delta|h|} dh$$

$$= \frac{F_r(1 - F_r)}{\pi} \frac{\lambda\Delta}{(\lambda\Delta)^2 + \varepsilon^2} . \qquad (5.6.12)$$

For $r \neq s$,

$$C_{rs}(h) = F_r(1 - e^{-\lambda\Delta|h|})F_s - F_r F_s$$

$$= -F_r F_s e^{-\lambda\Delta|h|}$$

so that

$$f_{rs}(\varepsilon) = \frac{-F_r F_s}{\pi} \frac{\lambda\Delta}{(\lambda\Delta)^2 + \varepsilon^2} . \qquad (5.6.14)$$

Noting that each $f_{rs}(\varepsilon)$ is continuous at $\varepsilon = 0$, and

$$tr[f(\varepsilon)] = \frac{\lambda\Delta}{[(\lambda\Delta)^2 + \varepsilon^2]} \sum_{r=1}^{\infty} F_r(1 - F_r),$$

which is bounded by tr[f(0)]. Therefore, all conditions have been satisfied, and

$$\sqrt{n}\begin{pmatrix} \frac{N_1}{n} - F_1 \\ \vdots \\ \frac{N_k}{n} - F_k \end{pmatrix}$$

has an asymptotic normal distribuiton with zero mean vector and asymptotic variance-covariance matrix, $V^{(n)}$, where

$$v_{jj}^{(n)} = \frac{2F_j(1 - F_j)}{\lambda\Delta}$$

and

$$v_{rs}^{(n)} = \frac{-2F_rF_s}{\lambda\Delta}$$

for $r \neq s$. To estimate the terms of $V^{(n)}$, recall from (5.4.2) that

$$\frac{M_s^j}{(n - s)F_j(1 - F_j)}$$

is an unbiased estimator of

$$1 - \pi(s\Delta) = 1 - e^{-\lambda s\Delta}$$

for a Poisson Lines Model. Using a Taylor expansion on $e^{-\lambda s\Delta}$ yields

$$\frac{\sum_{j=1}^{k} M_s^j}{(n - s)\left[1 - \sum_{j=1}^{k} F_j^2\right]}$$

as an estimator of $\lambda s\Delta$. Thus,

$$\frac{1}{n-1}\sum_{s=1}^{n-1}\left[\frac{\sum_{j=1}^{k} M_s^j}{(n-s)\left[1-\sum_{j=1}^{k} F_j^2\right]s\Delta}\right] \qquad (5.6.15)$$

is a reasonable estimator of λ. With N_j/n estimating F_j, $\mathrm{Var}\left(\frac{N_j}{n}\right)$ can now be estimated. Also, recalling from the lineal estimation procedure that

$$V^*\left(\frac{L_j}{L}\right) = \frac{2p_j(1-p_j)}{\lambda L}$$

for a lineal Poisson Process of intensity λ, the result from systematic point counts yields a similar expression with L being approximately equal to $n\Delta$, and F_j estimating p_j.

5.7 Extensions of Point Sampling

In developing the theory behind point sampling on the line, it was noted that the line represents a one-dimensional cut from a higher-dimensioned mosaic. Koop's 1976 results come from 21 different variations of point sampling in two dimensions. Two methods will be discussed, since their results can be tied in with the results already obtained in this chapter.

The first sampling procedure of random points on random lines locates b lines randomly and independently inside the mosaic, such that a points are placed at random locations on each line, accounting for $n = ab$ points on the plane. Define

$$u_{im}^{j} = \begin{cases} 1 & \text{if the } i^{th} \text{ point on the } m^{th} \text{ line belongs to a phase } j \text{ feature} \\ 0 & \text{otherwise,} \end{cases}$$

and let

p_{mj} = the proportion of the line occupied by the j-th phase.

and

p_j = the proportion of the plane occupied by the j-th phase.

Then

$$P(u_{im}^{j} = 1 \mid \text{line } m) = p_{mj},$$

and

$$E(u_{im}^{j}) = E_m E(u_{im}^{j} \mid \text{line } m) = p_j .$$

Thus,

$$\frac{N_j}{n} = \frac{\sum_{i=1}^{a} \sum_{m=1}^{b} u_{im}^{j}}{n}$$

gives an unbiased estimate of p_j, with

$$V\left(\frac{N_j}{n}\right) = \frac{1}{n^2}\left[\sum_{i=1}^{a}\left\{\sum_{m=1}^{b} \operatorname{var}(u_{im}^{j}) + \sum_{m\neq t} \operatorname{cov}(u_{im}^{j}, u_{it}^{j})\right\}\right.$$

$$\left. + \sum_{i\neq s}\left\{\sum_{m=1}^{b} \operatorname{cov}(u_{im}^{j}, u_{sm}^{j}) + \sum_{m\neq t} \operatorname{cov}(u_{im}^{j}, u_{st}^{j})\right\}\right] .$$

Due to the random and independent location of the lines on the plane,

$$\mathrm{cov}(u_{im}^{j}, u_{st}^{j}) = 0$$

for all $m \neq t$, so that only the first and third terms of the variance expression need to be considered. Now

$$V(u_{im}^{j}) = p_j(1-p_j),$$

and

$$\begin{aligned}
&\mathrm{cov}(u_{im}^{j}, u_{sm}^{j}) \\
&\quad = E_m \mathrm{cov}(u_{im}^{j}, u_{sm}^{j} \mid \text{line } m) + \underset{m}{\mathrm{cov}}\, E(u_{im}^{j}, u_{sm}^{j} \mid \text{line } m) \\
&\quad = \underset{m}{\mathrm{cov}}(p_{mj}, p_{mj}) = \mathrm{Var}(p_{mj}) = \sigma^2_{\frac{L_j}{L}},
\end{aligned}$$

where

$$\sigma^2_{\frac{L_j}{L}} = v^*_{jj}$$

as previously derived in Section 3.4.

Thus,

$$\begin{aligned}
V\left(\frac{N_j}{n}\right) &= \frac{p_j(1-p_j)}{n} + \frac{ab(b-1)}{n^2}\,\sigma^2_{\frac{L_j}{L}}, \\
&= \frac{p_j(1-p_j)}{n} + \left(\frac{1}{a} - \frac{1}{n}\right)\sigma^2_{\frac{L_j}{L}}, \qquad (5.7.1)
\end{aligned}$$

which yields the same result as Koop [9]. Note that the presence of $\sigma^2_{\frac{L_j}{L}}$ in the variance expression reflects the two-stage process of first sampling lines from the mosaic, then subsampling points from each transect. Note also that one needs a value for $\sigma^2_{\frac{L_j}{L}}$ even though point sampling is employed.

5.8 Systematic Points on Random Lines

The other sampling procedure to be discussed again begins with the independent location of b random lines inside the mosaic. On each line is placed a systematic array of a points, again resulting in $n = ab$ points on the plane. Using the same notation as in Section 5.7,

$$\frac{N_j}{n} = \frac{\sum_{i=1}^{a} \sum_{m=1}^{b} u^j_{im}}{n}$$

gives an unbiased estimate of p_j, with

$$V\left(\frac{N_j}{n}\right) = \frac{1}{n^2}\left[\sum_{i=1}^{a} \sum_{m=1}^{b} \mathrm{var}(u^j_{im}) + \sum_{i \neq s} \sum_{m=1}^{b} \mathrm{cov}(u^j_{im}, u^j_{sm})\right],$$

$$\begin{aligned} V(u^j_{im}) &= V_m E(u^j_{im} \mid \text{line } m) + E_m V(u^j_{im} \mid \text{line } m) \\ &= V_m(p_{mj}) + E_m[p_{mj}(1 - p_{mj})] \\ &= p_j(1 - p_j), \end{aligned}$$

and

$$\mathrm{cov}(u_{im}^{j},u_{sm}^{j})$$

$$= E_m \mathrm{cov}(u_{im}^{j},u_{sm}^{j} \mid \text{line } m) + \underset{m}{\mathrm{cov}}\, E(u_{im}^{j},u_{sm}^{j} \mid \text{line } m)$$

$$= E_m[p_{mj}(1-p_{mj})\pi(|i-s|\Delta) + \underset{m}{\mathrm{cov}}(p_{mj},p_{mj})]$$

$$= \pi(|i-s|\Delta)[p_j(1-p_j) - \sigma^2_{\frac{L_j}{L}}] + \sigma^2_{\frac{L_j}{L}} \ . \qquad (5.8.1)$$

Thus, the resulting expression for $V\left(\frac{N_j}{n}\right)$ will look like (5.3.2) with additional terms of $\sigma^2_{\frac{L_j}{L}}$ to account for the two-stage sampling procedure.

5.9 Comparisons with Hilliard-Cahn

Hilliard and Cahn's work with one-dimensional random point counts gives an expression that is the same as (5.7.1). Their results from systematic point counts are derived under very rigid restrictions, one being that a phase j feature will not occupy more than 1 point on the line. To equate this with the conditions described earlier in Section 5.3 would necessitate

(i) cells being extremely small or equivalently n being small and Δ being large in relation to L.

(ii) F_j being very small.

Under all of these restrictions, the Hilliard-Cahn variance expression becomes

$$V\left(\frac{N_j}{n}\right) = \frac{F_j}{n} ,$$

which would yield a very special case of the result from (5.3.2).

CHAPTER 6

MULTISTAGE SAMPLING

6.1 Introduction

As mentioned earlier, it is possible that the lineal and point sampling procedures discussed in Chapters 3, 4, and 5 may have resulted from a subsampling of a higher dimensional sectioning of the mosaic. When this occurs, multiple variance terms must be accounted for at each stage. The use of a lineal transect may yield estimates for structural features of a two-dimensional mosaic, but that planar mosaic may, in fact, be a cross-section of a three-dimensional figure. The volume-fraction of the void in a piece of copper might only be studied from looking at a cross-section of the specimen under a microscope. If lineal or point sampling is used, the sample fraction due to the void will in essence be estimating the areal fraction of the void on the planar sample. Thus, any results that would be obtained from a point count or lineal measurement may have to account for one or more additional stages of sampling.

6.2 Lineal Subsampling

The first case to be considered involves passing a lineal transect through a planar section of area A, which has been sampled from a higher-ordered mosaic. With notation as defined in Chapters 3 and 4, let L_j denote the total length of phase j segments on the line. Finding the expectation of L_j/L will now be a 2-step process, first conditioning on the planar section, $\mathcal{a}$, and then taking the expectation over the entire mosaic. As proven by Miles [15] in (2.2.4),

$$E\left(\frac{L_j}{L} \,\middle|\, \mathcal{a}\right) = \frac{A_j}{A} ,$$

and the expectation of A_j/A over the entire mosaic will be $V_j/V = p_j$.

To obtain the variance of L_j/L as both L and A go to infinity, the two-stage process uses the asymptotic variance approximations from Chapters 3 and 4. One has that

$$V^*\left(\frac{L_j}{L}\right) = V^*_{\mathcal{a}} E\left(\frac{L_j}{L} \,\middle|\, \mathcal{a}\right) + E_{\mathcal{a}} V^*\left(\frac{L_j}{L} \,\middle|\, \mathcal{a}\right)$$

$$= V^*_{\mathcal{a}}\left(\frac{A_j}{A}\right) + E_{\mathcal{a}}\left[\frac{1}{L\mu}(\sigma^2 + \mu^2)\,\frac{A_j}{A}\left(1 - \frac{A_j}{A}\right)\right] ,$$

where μ and σ^2 denote respectively the mean and variance of a cell intercept. Therefore,

$$V^*\left(\frac{L_j}{L}\right) = \sigma^2_{\frac{A_j}{A}} + \frac{1}{L\mu}(\sigma^2 + \mu^2)\left[p_j - E^2\left(\frac{A_j}{A}\right)\right]$$

$$= \sigma^2_{\frac{A_j}{A}} + \frac{1}{L\mu}(\sigma^2 + \mu^2)\left[p_j - p_j^2 - \sigma^2_{\frac{A_j}{A}}\right]$$

$$= \sigma^2_{\frac{A_j}{A}} + \sigma^2_{\frac{L_j}{L}} - \frac{\sigma^2_{\frac{L_j}{L}}\ \sigma^2_{\frac{A_j}{A}}}{p_j(1 - p_j)} \qquad (6.2.1)$$

for $p_j \neq 0,1$, where $\sigma^2_{\frac{L_j}{L}}$ and $\sigma^2_{\frac{A_j}{A}}$ denote the one-stage variance expressions from (3.4.4) and (4.2.1) respectively. An estimate for $\sigma^2_{\frac{L_j}{L}}$ can be made from the methods described in Chapter 3. However, information on the areal section will be needed to gather an estimate for $\sigma^2_{\frac{A_j}{A}}$.

6.3 Point Subsampling

If areal estimation is to be achieved through point sampling, the end result may be attributed to subsampling from a lineal transect, as developed by Koop [9] and mentioned in Chapter 5. Regardless of whether a random or systematic point cound is used, N_j/n as defined in Chapter 5 will be an unbiased estimator for A_j/A. For a two-stage systematic point count on a line $\mathcal{L}$, the asymptotic variance,

$$V^*\left(\frac{N_j}{n}\right) = V^*_{\mathcal{L}}E\left(\frac{N_j}{n}\,\middle|\,\text{line }\mathcal{L}\right) + E_{\mathcal{L}}V^*\left(\frac{N_j}{n}\,\middle|\,\mathcal{L}\right)$$

$$= \sigma^2_{\frac{L_j}{L}} + E_{\mathcal{L}}\left[\frac{\frac{L_j}{L}\left(1 - \frac{L_j}{L}\right)}{n}\left(1 + \frac{2}{n}\sum_{m=1}^{n}(n - m)\pi(m\Delta)\right)\right],$$

which reduces to

$$V^*\left(\frac{N_j}{n}\right) = \sigma^2_{\frac{L_j}{L}} + \sigma^2_{\frac{N_j}{n}} - \frac{\sigma^2_{\frac{L_j}{L}}\,\sigma^2_{\frac{N_j}{n}}}{\left(\frac{A_j}{A}\right)\left(1 - \frac{A_j}{A}\right)}\,, \qquad (6.3.1)$$

with approximations as developed earlier in Chapters 3 and 5.

If the systematic point process is generated from a lineal transect which in turn has been taken from an areal section of the mosaic, then a three-stage sampling has occurred, and will involve variance components from all three stages. Again, N_j/n is unbiased for p_j, with

$$V^*\left(\frac{N_j}{n}\right) = \underset{\mathcal{G},\mathcal{L}}{\text{Var}}\; E\left(\frac{N_j}{n} \,\middle|\, \text{line } \mathcal{L}\right) + \underset{\mathcal{G},\mathcal{L}}{E}\; V^*\left(\frac{N_j}{n} \,\middle|\, \mathcal{L}\right)$$

$$= \underset{\mathcal{G},\mathcal{L}}{V^*}\left(\frac{L_j}{L}\right) + E_{\mathcal{G},\mathcal{L}}\left\{\left[\frac{\frac{L_j}{L}\left(1 - \frac{L_j}{L}\right)}{n}\right]\left[1 + \frac{2}{n}\sum_{m=1}^{2}(n-m)\pi(m\Delta)\right]\right\}$$

$$= \sigma^2_{\frac{L_j}{L}} + \sigma^2_{\frac{A_j}{A}} - \frac{\sigma^2_{\frac{A_j}{A}}\,\sigma^2_{\frac{L_j}{L}}}{p_j(1 - p_j)} + \left[\frac{1}{n} + \frac{2}{n^2}\sum_{m=1}^{n}(n-m)\pi(m\Delta)\right]$$

$$\left[p_j - p_j^2 - \sigma^2_{\frac{A_j}{A}} - \sigma^2_{\frac{L_j}{L}} + \frac{\sigma^2_{\frac{A_j}{A}}\,\sigma^2_{\frac{L_j}{L}}}{p_j(1 - p_j)}\right]$$

$$= \frac{\sigma^2_{A_j}}{A} + \frac{\sigma^2_{L_j}}{L} + \frac{\sigma^2_{N_j}}{n} - \frac{1}{p_j(1 - p_j)}\left[\frac{\sigma^2_{A_j}}{A}\,\frac{\sigma^2_{L_j}}{L} + \frac{\sigma^2_{A_j}}{A}\,\frac{\sigma^2_{N_j}}{n} + \frac{\sigma^2_{L_j}}{L}\,\frac{\sigma^2_{N_j}}{n}\right]$$

$$+ \frac{\dfrac{\sigma^2_{N_j}}{n}\,\dfrac{\sigma^2_{L_j}}{L}\,\dfrac{\sigma^2_{A_j}}{A}}{[p_j(1 - p_j)]^2} \quad . \qquad (6.5.2)$$

Note that the three-stage sampling procedure incorporates the asymptotic variances from all three stages in a recognizable pattern, parallel to the results achieved by Anderssen [1].

CHAPTER 7

NUMERICAL RESULTS

7.1 Introduction

The variance approximations suggested in the past four chapters have been individually developed under specific assumptions on the mosaic. This chapter will numerically compare some of the approximations developed from Chapter 3, so that some practical guidelines can be given on the variance expression to be used for estimates obtained from lineal sampling on the mosaic. Four mosaics were each sampled by placing a grid of lines of fixed length within the boundaries of the two-dimensional pictures. Table I gives the results of the lineal sampling of a random lines mosaic (Poisson Lines Model) given by Pielou [16]. The mosaic studied by Table II is a simple case from Matérn [12] of a realization of the Boolean Scheme, with each β_j set being a circle of fixed radius. The results of Table III come from an ad hoc picture consisting of many small features, which are similar in size and shape, and fairly regularly spaced. Tables I - III are used to compare the variance expressions of the areal-fraction estimators. Table IV gives variance approximations for data gathered to estimate the boundary length of Lake Alice, on the University of Florida campus.

The areal-fraction variance terms compared in Tables I - III are the v_{jj}^* terms under:

(1) The assumption of the Poisson Lines Model, where v_{jj}^* is given by (3.11.1).

(2) The assumption of the Boolean Scheme, which gives v_{jj}^* by (3.15.4).

(3) The technique developed by Hilliard and Cahn [8], which utilizes the individual segment information by (3.10.1).

7.2 Results: Table I

The results of Table I give very reasonable approximations from the expression (3.11.1), in comparison with

$$s^2_{\frac{L_j}{L}} = .0426,$$

which is the sample variance of the individual areal-fraction estimates from the L_j/L column. The only times for which the approximation seemed to differ greatly was for lines 3 and 4, for which L_j/L was small, resulting in the numerator of v_{11}^* being small.

The Hilliard-Cahn approximations from these lines tended to underestimate the lineal-fraction variance. This was not a surprising result, in that the model, obtained from Pielou [16], resulted in cell sizes that were fairly large in comparison with the length of the sampling transect. This allowed for very few segments from each phase, and accounted for minimal variability from segment to segment, which results in a small ratio of the s_j^2 terms to the $\bar{y}_j^2$ terms. Note that in line 7, only one phase 2 segment was on the transect, so that an approximation by this approach was not possible.

In looking at the results over all ten lines, it was suggested that a wider spacing of the lines might produce a more "random" sampling effect, so the second section of Table I was added. This resulted in approximations that were very close to the sample variance terms of the lineal-fraction estimates.

7.3 Results: Table II

Originally, the mosaic yielding the results of Table II was sampled in order to test the variance approximations from a mosaic arising from a Boolean Scheme. Unfortunately, the picture taken from Matérn [12] again had large β_j sets in comparison with the fixed length of the sampling transect. This resulted in individual variance approximations using (3.15.4) being very close to zero, and sometimes negative. In looking at (3.15.4),

$$\hat{v}_{jj}^{*} = \frac{\left(\frac{L_j}{L}\right)\left(1 - \frac{L_j}{L}\right)(4r)}{L} - \frac{2r^2M}{L^2} .$$

The first term of this expression yielded values that were close to $s_{\frac{L_j}{L}}^2 = .0301$, but the second term tended to subtract off too much, resulting in the variance approximation being too small. This might be accounted for from only generating the first two terms in the Taylor expansion for $e^{\tau K(h)}$, in obtaining (3.15.4). The third term might be able to compensate for the amount subtracted by the second term.

This model did allow for comparisons between the Hilliard-Cahn and Poisson Lines approximations for the variance of L_1/L. Again, the Hilliard-Cahn approximations tended to be smaller than the sample variance, which can be attributed to the regularity of the β_j sets (circles of same radius), thus allowing for a minimal amount of variability from segment to segment.

The approximations under the Poisson Lines Model were much more accurate than the Hilliard-Cahn results, although in general, they also tended to underestimate $s^2_{\frac{L_j}{L}} = .0301$.

It was interesting to see how the assumption of an incorrect model would affect the variance approximations of the lineal-fraction estimates. For this case, the approximations derived from the Poisson Lines Model used on a Boolean Scheme mosaic were still fairly good.

Again, by looking at a wider spacing of the lines, it was hoped that the estimate from the assumption of the Boolean Scheme might give some accurate variance approximations, but again the second term of (3.15.4) tended to subtract too much from the variance expression. However, the Poisson Lines approximations seemed to be reasonably accurate with the wider line spacing.

7.4 Results: Table III

After studying the results from Tables I and II, it was decided to try to find a model for which the Hilliard-Cahn technique might be appropriate. The mosaic studied by Table III had many small features with relatively little variation, so that each sampling transect studied had four phase one segments and five phase two segments. With this much information to utilize, the Hilliard-Cahn technique gave reasonable approximations to the sample variance. Five additional lines were included in the sampling grid to check for variability of the areal fraction estimates, and the sample variance from the 10 line grid made the Hilliard-Cahn results look even more accurate than ever before, and certainly much improved over the results from Tables I and II.

Surprisingly enough, again under false modeling assumptions, the Poisson Lines variance approximations seemed to be fairly close to the ten-line sample variance. However, this was the first time in the three models studied that these approximations tended to generally exceed the sample variance.

7.5 Results: Table IV

The sampling from the University of Florida map was originally done for the purpose of estimating the boundary length of Lake Alice. Recall from (2.2.6) that the $\pi/2$ times the number of intersections per unit length the sampling transect makes with the boundary gives an unbiased estimate of boundary length per unit area. Modifying the variance term of (3.8.7),

$$V^{*}\left(\frac{N_1}{L}\right) = \frac{4}{L}\left[C_1\{p_1^2 + (1-p_1)^2\} + C_1^2\left\{\frac{\mu^{(2)}}{\mu} - 2\mu\right\}\right] . \qquad (7.5.1)$$

Estimating C_1 by $N_1/2L$ and p_1 by L_1/L gives

$$\hat{V}^{*}\left(\frac{N_1}{L}\right) = \frac{2N_1}{L^2}\left[\left(\frac{L_1}{L}\right)^2 + \left(1 - \frac{L_1}{L}\right)^2\right] , \qquad (7.5.2)$$

assuming the C_j^2 term of (7.5.1) will be very small. From the four line grid used on Lake Alice, the variance terms tended to exceed the sample variance of the N_1/L values, but when a grid of 15 lines was used on the mosaic, the sample variance of N_1/L was bigger than any of the individual variance approximations, implying that this variance approximation appears to be reasonable for boundary length estimation.

7.6 Conclusions

While no major results could be concluded from a study of three theoretical mosaics, several observations can be made from the results of the four tables. For any variance approximation to yield accurate results, the features should be small in relation to the overall size of the mosaic, which could allow for a relatively long sampling transect. The results from the Hilliard-Cahn technique appear to be more and more accurate when the individual features are widely varied, and the transect intersects many features. The Poisson Lines approximations appear to be reasonable for a random arrangement of features, which can be tested from the method described in Section 3.13. However, the Poisson Lines approximations also gave close results under the false modeling assumptions, so that this method might be used for situations in which sampling information is not easily obtainable.

Unfortunately, the Boolean Scheme only gave crude approximations, for reasons explained in Section 7.3, but perhaps for models with very small, but fairly congruent features randomly situated about the mosaic, this method might yield significant results, even if only the first term of (3.15.4) is used.

The variance approximation from Section 7.5, used in boundary length estimation seemed to be reasonably close to the sample variance, and may yield even better results for irregularly shaped features that may not necessarily be convex.

Table I

Line	L_1/L	L_2/L	M_1	M_2	Poisson $\hat{v}^*_{11}$	Hilliard-Cahn $\hat{v}^*_{11}$
1	.316	.684	2	2	.0467	.0274
2	.337	.663	2	2	.0499	.0112
3	.168	.832	2	3	.0156	.0163
4	.200	.800	2	3	.0205	.0895
5	.337	.663	2	3	.0399	.0139
6	.474	.526	2	2	.0622	.0237
7	.737	.263	2	1	.0501	---
8	.716	.284	2	2	.0413	.0393
9	.663	.337	3	2	.0399	.0208
10	.453	.547	3	3	.0491	.0593
Averages	.4401	.5599	2.2	2.3		

Over all 10 lines:

Poisson: $\hat{V}(\hat{p}_1) = .0552$ $\quad s^2_{\frac{L_j}{L}} = .0426$

Hilliard-Cahn: $\hat{V}(\hat{p}_1) = .0264$

Lines	$\hat{p}_1$	$s^2_{\hat{p}_1}$	Poisson $\hat{V}(\hat{p}_1)$
1,3,5,7,9	.4442	.0595	.0544
2,4,6,8,10	.4360	.0364	.0550
1,4,7,10	.4265	.0536	.0563
2,5,8	.4633	.0480	.0571
3,6,9	.4350	.0624	.0519

Table II

Line	L_1/L	L_2/L	M_1	M_2	Poisson $\hat{v}^*_{11}$	Hilliard Cahn $\hat{v}^*_{11}$	Boolean Scheme (1st term only) $\hat{v}^*_{11}$
1	.191	.809	2	3	.0191	.0122	.0355
2	.124	.876	2	2	.0118	.0098	.0248
3	.514	.486	5	5	.0250	.0164	.0571
4	.629	.371	5	5	.0218	.0107	.0533
5	.514	.486	5	5	.0250	.0218	.0571
6	.448	.552	5	5	.0245	.0082	.0565
7	.314	.686	4	3	.0265	.0102	.0492
8	.276	.724	3	4	.0228	.0062	.0457
9	.219	.781	2	3	.0234	.2000	.0391
Averages	.3588	.6412	3.67	3.89			

Over all 9 lines:

Poisson: $\hat{V}(\hat{p}_1) = .0280$ $s^2_{\frac{L_j}{L}} = .0301$

Hilliard-Cahn: $\hat{V}(\hat{p}_1) = .0216$

Boolean Scheme: $\hat{V}(\hat{p}_1) = .0098$

Boolean Scheme (1st term only): $\hat{V}(\hat{p}_1) = .0526$

Lines	$\hat{p}$	$s^2_{\hat{p}}$	Poisson $\hat{V}(\hat{p}_1)$	Boolean Scheme $\hat{V}(\hat{p}_1)$	Boolean Scheme (1st term only) $\hat{V}(\hat{p}_1)$
1,2,5,7,9	.3504	.0244	.0280	.0102	.0520
2,4,6,8	.3693	.0475	.0280	.0091	.0532
1,4,7	.3780	.0510	.0301	.0123	.0537
2,5,8	.3047	.0387	.0257	.0088	.0484
3,6,9	.3937	.0240	.0274	.0067	.0546

Table III

Line	L_1/L	L_2/L	M_1	M_2	Poisson $\hat{v}^*_{11}$	Hilliard-Cahn $\hat{v}^*_{11}$
1	.3353	.6647	4	5	.0221	.0125
2	.1824	.8176	4	5	.0099	.0089
3	.2235	.7765	4	5	.0134	.0135
4	.2294	.7706	4	5	.0139	.0047
5	.2647	.7353	4	5	.0168	.0060
Averages	.2471	.7529	4	5		

Over all 5 lines:

Poisson: $\hat{V}(\hat{p}_1) = .0154$ $s^2_{\frac{L_1}{L}} = .0032$

Hilliard-Cahn: $\hat{V}(\hat{p}_1) = .0093$

Over 10 lines:

$$s^2_{\frac{L_1}{L}} = .0108$$

Table IV

Line	L_1/L	N_1	$\hat{V}(N_1/L)$
1	.395	2	.000318
2	.235	2	.000390
3	.568	4	.000621
4	.469	4	.000612
Average	.4168	3	

Over all 4 lines:

$$\hat{V}(\bar{N}_1/L) = .000470 \qquad s^2_{\frac{N_1}{L}} = .000203$$

From a 15 line grid:

$$s^2_{\frac{N_1}{L}} = .001085$$

BIBLIOGRAPHY

[1] Anderssen, R. S. (1975). Recent Results in Volume-Fraction Analysis. Australian National University, Report V-FA.

[2] Billingsley, P. (1968). Convergence of Probability Measures. New York: Jonh Wiley and Sons.

[3] Cox, D. R. (1962). Renewal Theory. New York: John Wiley and Sons.

[4] Davy, P. and Miles, R. E. (1971). Sampling theory for opaque spatial specimens. JRSS-B, 39-1.

[5] DeHoff, R. T. and Rhines, F. N. (1968). Quantitative Microscopy. New York: McGraw-Hill.

[6] Gilbert, E. N. (1962). Random subdivisions of space into crystals. AMS, 33, 958 - 72.

[7] Hannan, E. J. (1970). Multiple Time Series. New York: John Wiley and Sons.

[8] Hilliard, J. E. and Cahn, J. W. (1961). An evaluation of procedures in quantitative metallography for volume-fraction analysis. Trans. Met. Soc. AIME, 221, 344 - 52.

[9] Koop, J. C. (1976). Systematic Sampling of Two-Dimensional Surfaces and Related Problems. Research Triangle Institute Project Number 250V-1123.

[10] Lang, A. R. G. and Melhuish, F. M. (1970). Lengths and diameters of plant roots in non-random populations by analysis of plane surfaces. Biometrics, 26, 421 - 31.

[11] Matérn, B. (1960). Spatial Variation: Stochastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations. Meddelanden Fran Statens Skogsforskningsintitute, Band 49, NR5.

[12] Matérn, B. (1972). Analysis of Ecological Maps as Mosaics. Talk given at The Advanced Institute on Statistical Ecology in the United States.

[13] Miles, R. E. (1964). Random polygons determined by random lines in a plane, I. Proc. Nat. Acad. Sci., USA, 52, 901 - 7.

[14] Miles, R. E. (1972). The random division of space. Adv. App. Prob., 4-4, 243 - 66.

[15] Miles, R. E. (1972). Multi-dimensional perspectives on Stereology. J. Microscopy, 95-II, 181 - 205.

[16] Pielou, E. C. (1967). An Introduction to Mathematical Ecology. New York: John Wiley and Sons.

[17] Pyke, R. (1961). Markov renewal processes: definitions and preliminary properties. AMS, 32, 1231 - 42.

[18] Pyke, R. and Schaufele, R. (1964). Limit Theorems for Markov renewal processes. AMS, 35, 1746 - 64.

[19] Scheaffer, R. L. (1973). Tests for uniform clustering and randomness. Comm. in Stat., 2(6), 479 - 92.

[20] Scheaffer, R. L. (1974). An Approximate Variance for Line Intersection Counts. Tech. Report 76, Department of Statistics, Univ. of Florida.

[21] Scheaffer, R. L. (1975). Variance Approximations for Transects: The Poisson Lines Model. Tech. Report 86, Department of Statistics, Univ. of Florida.

[22] Serra, J. The Boolean Scheme and Poisson Flats. Unpublished Notes on Mathematical Morphology.

[23] Smith, W. L. (1955). Regenerative stochastic processes. Proc. Royal Soc. London, A, 323, 6 - 31.

[24] Switzer, P. (1965). A random set process in the plane with a Markovian property. AMS, 36, 1859 - 63.

[25] Weibel, E. R. (1972). The value of stereology in analysing structural function of cells and organs. J. Microscopy, 95-I, 3 - 13.

BIOGRAPHICAL SKETCH

James Robert Lackritz was born December 30, 1950, in Columbus, Ohio, where he spent his first fourteen years of his life before moving to Shaker Heights, Ohio. Upon graduation from Shaker Heights High School in 1968, he enrolled at Bucknell University, receiving the Bachelor of Science Degree in Mathematics in May, 1972. While at Bucknell, he developed an interest in statistics under the tutelage of the late Professor Paul Benson. In September, 1972, James began his graduate studies at the University of Florida, and received a Master of Statistics degree in June, 1974. He has been teaching undergraduate courses at the University of Florida while pursuing the degree of Doctor of Philosophy.

After receiving his degree, James will join the faculty at San Diego State University as an Assistant Professor in the Department of Management.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Richard L. Scheaffer, Chairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James T. McClave

James T. McClave
Associate Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

P. V. Rao

P. V. Rao
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

F. H. Dietrich II

F. H. Dietrich, II
Assistant Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully acceptable, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Robert T. DeHoff

Robert T. DeHoff
Professor of Materials Science
Engineering

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Arts and Sciences and to the Graduate Council, and was accepted as partial fulfillment of the Requirements for the degree of Doctor of Philosophy.

August 1977

Dean, Graduate School

UNIVERSITY OF FLORIDA
3 1262 08666 278 9